12-2007

Applied Fourier Transform Near-infrared Techniques for Biomass Compositional Analysis

Lu Liu
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Lu Liu entitled "Applied Fourier Transform Near-infrared Techniques for Biomass Compositional Analysis." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Biosystems Engineering.

X. Philip Ye, Major Professor

We have read this thesis and recommend its acceptance:

Douglas G. Hayes, Alvin R. Womac, Arnold M. Saxton

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Lu Liu entitled “Applied Fourier Transform Near-infrared Techniques for Biomass Compositional Analysis.” I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Biosystems Engineering.

X. Philip Ye, Major Professor

We have read this thesis
and recommend its acceptance:

Douglas G. Hayes

Alvin R. Womac

Arnold M. Saxton

Accepted for the Council:

Carolyn R. Hodges, Vice Provost
and Dean of the Graduate School

(Original signatures are on file with official student records.)

Applied Fourier Transform Near-Infrared Techniques for Biomass Compositional Analysis

A Thesis Presented for
the Master of Science
Degree
The University of Tennessee, Knoxville

Lu Liu
December 2007

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. X. Philip Ye, for giving me this great opportunity in this biomass project, and teaching me how to be a good researcher. I appreciate all the advice and support from my committee members: Drs. Al. Womac, Douglas Hayes and Arnold Saxton. Thank you all for serving on my committee.

I would also like to send my appreciation to many colleagues: Margaret Taylor, Lindsey Kline, Galina Melnichenko, Haizhou Li, Brian Chomicki for their help with either my research or my English. Thanks to two undergraduates John Kruckeberg and Abdoulaye Samba for their assistance in the HHV measurement. Furthermore, I want to say thanks to my best friend Javier Gomez del Rio. He gave me a lot of advice in both my personal life and research, and this friendship accompanies me ever since I arrived in Knoxville.

Finally, I want to say thanks to my family. All their love and support helped me all the way through.

ABSTRACT

A new method for rapid chemical analysis of lignocellulosic biomass was developed using Fourier transform near-infrared (FT-NIR) spectroscopic techniques. The new method is less time-consuming and expensive than traditional wet chemistry. A mathematical model correlated FT-NIR spectra with concentrations determined by wet chemistry. Chemical compositions of corn stover and switchgrass were evaluated in terms of glucose, xylose, galactose, arabinose, mannose, lignin, and ash. Model development evaluated multivariate regressions, spectral transform algorithms, and spectral pretreatments and selected partial least squares regression, log(1/R), and extended multiplicative signal correction, respectively. Chemical composition results indicated greater variability in corn stover than switchgrass, especially among botanic parts. Also, glucose percentage was higher in internodes (>40%) than nodes or leaves (~30-40%). Leaves had the highest percentage of lignin (~23-25%) and ash (~4-9%). Husk had the highest total sugar percentage (~77%). Individual FT-NIR predictive models were developed with good accuracy for corn stover and switchgrass. Root mean square errors for prediction (RMSEPs) from cross-validation for glucose, xylose, galactose, arabinose, mannose, lignin and ash were 0.633, 0.620, 0.235, 0.374, 0.203, 0.458 and 0.266 (%w/w), respectively for switchgrass, and 1.407, 1.346, 0.201, 0.341, 0.321, 1.087 and 0.700 (%w/w), respectively for corn stover. A unique general model for corn stover and switchgrass was developed and validated for general biomass using a combination of independent samples of corn stover, switchgrass and wheat straw. RMSEPs of this general model using cross-validation were 1.153, 1.208, 0.425, 0.578, 0.282, 1.347 and 0.530 %w/w for glucose, xylose, galactose, arabinose, mannose, lignin and ash, respectively. RMSEPs for independent validation were less than those obtained by cross-validation. Prediction of major constituents satisfied standardized quality control criteria established by the American Association of Cereal Chemists. Also, FT-NIR analysis predicted

higher heating value (HHV) with a RMSEP of 53.231 J/g and correlation of 0.971. An application of the developed method is the rapid analysis of the chemical composition of biomass feedstocks to enable improved targeting of plant botanic components to conversion processes including, but not limited to, fermentation and gasification.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I INTRODUCTION

1.1 Biomass properties and industrial potentials for bio-energy

Broadly, biomass is all plant and plant-derived matter. For renewable energy resources, biomass often refers to agricultural residuals, wood waste, and dedicated energy crops. As fossil fuels are facing depletion (Hoel and Kverndokk, 1996), lignocellulosic biomass has attracted growing attention as a promising alternative. Various products can be derived from lignocellulosic biomass through thermal, chemical, biological, and physical conversions (Demirba ，2001; Kucuk, 1997). Biomass is a renewable energy source that will provide environmental and economic benefits.

Lignocellulosic biomass consists of three major constituents: cellulose, hemicellulose, and lignin (Gary et al., 1983). Other than serving different biological functions, these constituents have different industrial applications. Biomass materials with higher cellulose content are preferable for bio-ethanol production, while lignin-enriched biomass is preferred by combustion or co-firing with coal, since lignin has high heating value. Lignin can also be used in many other productions such as plastic products and polyblends (Hu, 2002). Ash has few applications. But knowledge of the ash content contributes to the selection of feedstock, enhances heat transfer and reduces the ash slagging and fouling problems in combustion (Winegartner, 1974, Baxter, 1993). The assessment of biomass chemical composition as well as the heating value aids feedstock selection and process adjustment for higher yield. Therefore, the contents of carbohydrates, and lignin, ash, as well as higher heating value (HHV) may be very important to the bio-energy industry, and were thus selected as the target analytes in this study.

1.2 The disadvantage of wet chemistry analyses

Total acid hydrolysis (ASTM E1758-95, 1995; Sluiter et al., 2006) is a traditional wet chemistry method for the structural carbohydrates and lignin measurement. A two-step acid hydrolysis is utilized to break down large

polymers into smaller detectable compounds. Many chemicals (sulfuric acid, calcium carbonate, high purity monosaccharides, etc) are consumed and various instruments (eg. analytical balance, autoclave, oven, HPLC, UV-Vis, etc) are involved. Ash is measured gravimetrically after 550~600 ºC (ASTM E1755-01) combustion. Heating value is measured using calorimeter and consumption of oxygen.

The wet chemistry approach has disadvantages that make it impractical, or even impossible, to be applied to at-line or online monitoring in the industry. First of all, the time-consuming characteristic is the major drawback. For instance, it takes over two days to obtain data on sugars, lignin, and ash. Second, all these conventional methods consume chemicals and energy, which increase analysis expense. It was reported that a complete analysis set via standard wet chemistry methods costs $800-2000 per sample (Hames et al., 2003). Third, operating wet chemistry analysis requires systematic laboratory settings, including the water purification system, filtration settings, analytical balance, autoclave, water bath, convection oven, furnace, desiccator, HPLC, UV-Vis, etc, and a variety of glassware as well. In addition, the wet chemistry method is labor-intensive and the operation steps such as weight measurements and filtrations may be the potential error sources, which affect the analysis reproducibility and precision.

Therefore, the traditional analysis methods for the chemical composition as well as heating value measurement will not satisfy the booming bio-energy industry. The time and expense issues make it difficult to expend the biomass compositional analyses via wet chemistry methods from laboratory scale to industrial scale. Researchers and engineers are trying every possibility in seeking an alternative analysis pathway.

1.3 An effective solution: near infrared technique with chemometrics

Near infrared (NIR) spectroscopy combined with multivariate analyses has been studied as an effective approach to chemical property prediction (Burn, 1997; Workman, 1999). The chemical bond of an object can be excited to higher energy level by absorbing the radiation at certain frequencies (*f*), which is depicted (Equation 1) in terms of wavenumber (υ) in cm^{-1} or wavelength (λ) in nm in an electromagnetic spectrum.

Equation 1: $\lambda = c/f = 1/\upsilon$ where: *c* is the speed of light

The electromagnetic spectrum is the range of all possible electromagnetic radiation, while near-infrared refers to the electromagnetic spectrum ranging from 4000 cm^{-1}~12500 cm^{-1} (800 nm to 2500 nm), which contains combination bands, 1st and 2nd overtones. Since the biomass is basically composed of organic compounds with various chemical bonds and the NIR spectrum is the accumulative result attributed to existing chemical bonds, mathematical models can be derived to correlate chemical information with NIR spectral information. Since the acquisition of biomass NIR spectra is much faster and simpler than wet chemistry analysis, the goal is to predict the chemical information based solely on the NIR spectra. In model calibration, independent variables come from NIR spectral data and the dependent variables come from chemical concentration data. With the developed model, the analyte content can be calculated from the sample spectrum, which can be acquired within minutes. This non-destructive approach does not involve chemical reactions, and thus minimizes expense. NIR analysis approach is promising for application in bio-energy industry, especially in online monitoring.

Efforts have been made in this NIR technique with chemometrics in many disciplines, such as forage (eg. Martens, et al., 1984; Melchhinger, et al., 1990; Flores Pires, et al., 1998; Xiccato et al., 1999), food (eg. Hong et al., 1996;

Manley et al., 2002; Maertens et al., 2004), and forestry (eg. Michell and Schimlec, 1994; Schimleck et al., 2000; Kelley, 2004; Lestander and Rhen, 2005; Balleirini dos Santos, 2006). Especially, the applications in wood and forestry areas indicate the feasibility of applying the NIR technique to biomass. However, although wood material and agricultural biomass are similarly composed, it does not mean that those available models for woods can be taken directly for use on agricultural biomass materials. Given that their constituents cover different ranges from biomass (Table 1), it is unreliable for the models developed for woods to predict agricultural residues. Even within the same range, due to the physical interferences, the feasibility of using the model of one species to predict another species needs verification.

1.4 Further development of NIR instrumentation—Fourier-transform (FT)-NIR

There are two common NIR spectrometers: dispersive NIR and FT-NIR nowadays. Dispersive NIR employs a prism or moving grating to separate each frequency that passes individually through a slit to the detector. The slit determines the spectral bandwidth (resolution) while the grating movement controls the passing frequency. Then the detector measures once a time the radiation at each frequency that has reflected from (or transmitted through) the sample (Koenig, 2001). Therefore, it takes time to complete all the measurement across the NIR spectral region. The major difference of FT-NIR is the interferometer system. The mechanism is illustrated in Figure 1. NIR beam from the source is split by the beamsplitter into two sub-fluxes. One sub-flux is sent to a stationary mirror, and another sub-flux goes to a moving mirror, which moves back and forth at a constant speed. The two sub-fluxes are reflected by the two mirrors respectively and merge at the beamsplitter with different traveling distance. In this way, some of the frequencies recombine constructively and some combine destructively and thus interferogram is formed.

Table 1 General chemical composition (%w/w) of different biological materials (DOE, 2004)

Material	*Six-Carbon Sugars*	*Five-Carbon Sugars*	*Lignin*	*Ash*
Hardwoods	39-50%	18-28%	15-28%	0.3-1.0%
Softwoods	41-57%	8-12%	24-27%	0.1-0.4%
Ag Residues	30-42%	12-39%	11-29%	2-18%

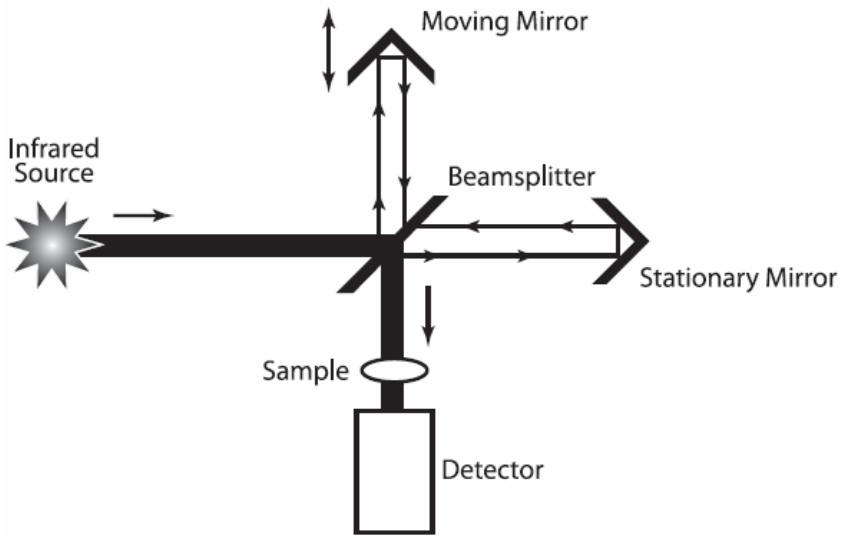


Figure 1 Illustration of FT-NIR spectrometer layout

(Thermo application note TN-00128, 2002)

This interferogram then goes to the sample, and the reflected (or transmitted) portion is received by the detector with the entire NIR frequencies. In other words, the detector obtains complete information of every frequency simultaneously.

Different mechanical layouts determined numerous advantages of FT-NIR over dispersive NIR. First, FT-NIR measures all the frequencies simultaneously while dispersive NIR does it individually. So the scan time is much shorter for FT-NIR than the dispersive NIR; or given the same amount of time, FT-NIR can generate better spectral representation by averaging more scans (McCarthy and Kemeny, 2001). Second, the slit in dispersive NIR restricts the throughput energy; the higher resolution, the narrower the silt and thus the more substantial energy loss, which results in poor spectral quality. In contrast, there is no degradation of optical throughput for FT-NIR, so higher optical resolution is achieved by FT-NIR without compromising signal-to-noise ratio (Griffiths and de Haseth, 1986). Third, FT-NIR uses a laser to control the speed of the moving mirror and to provide internal wavelength calibration. Since the wavelength of the laser is a constant, the operation and calibration is precise and accurate by referring to this value. Therefore, FT-NIR has greater repeatability compared to dispersive NIR which relies on external calibration. In addition, the relatively mechanically simple apparatus conformation (less moving parts) ensures FT-NIR less possibility of breakdown (Griffiths and de Haseth, 1986).

Before the advent of FT-NIR instrumentation, most of NIR studies were conducted on agricultural and wood materials using dispersive NIR spectrometer, while there is still a deficiency in the studies using FT-NIR, especially on the subjects of biomass. So there is a need to enhance the NIR research with this more advanced equipment.

CHAPTER II BACKGROUND INFORMATION

2.1 Biomass feedstocks investigated in this study

2.1.1 Corn stover

Corn (*Zea mays L.*) stover is the major field crop residue in the US with an annual availability of over 238 million tons (Sokhansanj et al., 2002). The United States is the single largest corn producer in the world, with approximately 28 million hectares and taking up slightly less than 1/4 of all US cropland; almost 20% of the corn production is diverted to produce ethanol (Patzek, 2004) and corn is the major ethanol production feedstock according to USDA's 2002 survey (Shapouri and Gallagher. 2005). Since corn is a primary crop product, key food source and the feedstock for current bio-ethanol production, there is a huge resource of corn stover. Consequently, corn stover is considered in the near future as a renewable energy feedstock.

Corn stover has been studied using dispersive NIR spectroscopy (Hames et al., 2003), which showed the feasibility of NIR techniques on biomass chemical composition analysis. Therefore, this species is a good subject to start to further update the NIR analysis on corn stover with FT-NIR instrument. The available composition data resource of corn stover can be used as reference to the wet chemistry measurement in this study

2.1.2 Switchgrass

Switchgrass (*Panicum virgatum*) is one of the dominant tall grass species, perennially growing on the central North American prairie (Towne and Owensby, 1984). Switchgrass has been discussed frequently as major energy crop in recent years. Switchgrass was determined to be focused on by Department of Energy Bio-energy Feedstock Development Program after more than 30 herbaceous crops species had been screened, due to its high yield, excellent conservation attributes and good compatibility with conventional farming practices (McLaughlin et al., 1992).

Studies on switchgrass have been ongoing in studying their yields, energy potential, and environmental impact (McLaughlin et al., 1992; Hopkins et al., 1995; Christian et al., 1998; Elbersen et al., 1998; Duffy and Nanhou, 2002; Lemus et al., 2002; Boateng et al., 2006). All these research activities suggest that switchgrass is a promising feedstock for bio-fuel industry. However, few NIR studies have been done on switchgrass; even the compositional information, with respect to the renewable energy consideration, is insufficient for the switchgrass cultivars and the botanic parts. Therefore, this study will provide not only the switchgrass compositional information to further broaden current knowledge on switchgrass, but also a rapid compositional analysis approach using FT-NIR techniques.

2.2 Diffuse reflectance in NIR region

Diffuse reflectance mode is usually utilized in NIR analysis of solid or powder samples (Osborne, 1981). When incident NIR beam sheds on the sample particles, the radiation is distributed into several sub-flux (Figure 2): 1) some radiation is reflected directly off the surface; 2) some is directly absorbed by the surface particles; 3) some is reflected to the next particles; 4) some is transmitted through the surface particles to the inner particles; and also 5) the absorption, transmission and reflection may cross-occur many times in the

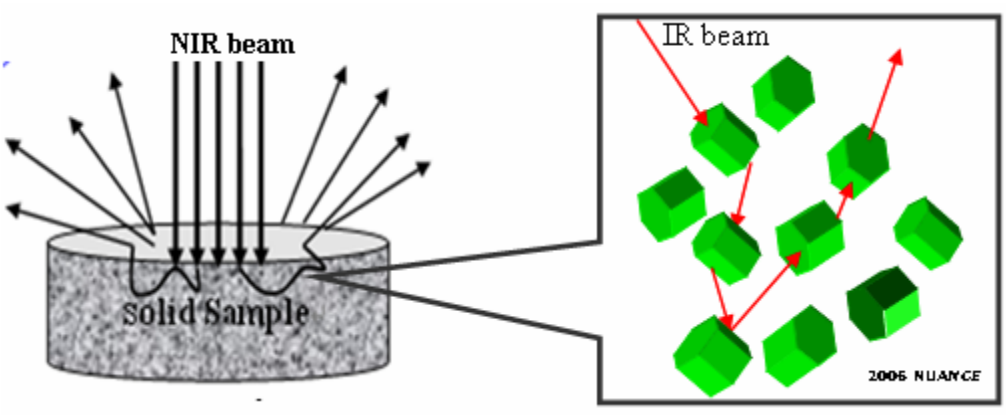


Figure 2 Illustration of NIR diffuse reflectance

sample until the remaining radiation reflected out of the surface. Amongst all these sub-fluxes, only those scattered within the sample and returned to the surface are considered as diffuse reflectance (Osborne, 1981, Bajcsy, 1996; Pasikatan et al., 2001).

The detector collected all the diffused reflected flux of the sample (I_d) and compared to that of a non-absorbing standard (I_o), resulting in the reflectance ($R = I_d/I_o$). The amount of radiation reflected from the sample is quantified as the reflectance (R) of the sample. As for diffuse reflectance NIR spectra, the intensity is commonly presented in two forms: *log(1/R)* and Kubelka-Munk (*K-M)*. The algorithm for *log(1/R)* is the logarithm of the inversed reflectance, while *K-M* is defined as Equation 2.

Equation 2

$$K\text{-}M = {(1-R)^2}\Big/{2R}$$

Where: R is the reflectance.

The detected NIR radiation was partially absorbed by the sample, thereby carrying the sample chemical information. This is the communication key between NIR spectra and the sample chemical characteristics. The concentrations of analytes can be statistically correlated to the spectral dataset (Kortum, 1969; Martens and Naes, 1987; Workman and Brown, 1996; Wold and Sjostrom, 1998), since these chemical concentration, or chemical bonds concentration were the essential cause of the NIR spectra. The major task in the model calibration is to seek a set of coefficients (K) via statistical and mathematical approaches to satisfy Equation 3.

Equation 3

$$C = K \cdot X$$

where C represents the concentration information of the chemical data, and X bears the spectral information.

Studies have proved that this mathematical equation is quantitatively and statistically derivable through multivariate analysis; however, the specific application varies. First, the spectral data (X) can have units either in *K-M* or

log(1/R). Second, the spectral data (X) can either represent the original wavenumber or represent the recomposed factors arising from the original scales. Third, even the concentration data can have different presentations, either in the original scales or the recomposed factors (Kramer, 1998).

2.3 Data processing methodology

2.3.1 Data pretreatment methods

When NIR techniques are in use to assess chemical composition information, the ideal scenario occurs when the differences among spectra are exclusively attributed to the chemical properties. However, many physical factors exist, especially in diffuse reflectance sampling, interfering with NIR spectral acquisition of chemical information. Many pretreatment algorithms are dedicated to correcting for the physical interferences. The most commonly used pretreatment methods are: standard normal variate (SNV), 1st derivative, 2nd derivative, multiplicative scatter correction (MSC), extended multiplicative signal correction (EMSC), as well as various combinations of them. The mechanisms of these pretreatment methods are described below.

Assume that w_i represents each wavenumber throughout the investigated NIR region, and y_i is the overall intensity shown along spectral ordinate corresponding to each w_i. Since NIR spectra combine information from both physical phenomena and chemical variation, y_i is not solely attributed to the chemical factors. Another variable x_i is necessary to represent the part of spectral signal corresponding solely to chemical information and the remaining part of y_i is then considered as all the possible additive variation introduced by the physical interference.

If the latent physical interference causes the multiplicative effect and offset, which thus modify the presented spectral to:

Equation 4 $$y_i = a + bx_i$$

Then SNV (Sánchez et al., 2004; Esbensen, 2004) could be a solution. Since

SNV is calculated as Equation 5, it centers and scales individual spectra without relying on the other spectra.

Equation 5 $$x_{SNV} = (y_i - \mu) / \sigma$$

Where: μ is the mean and σ is standard error

If the additive terms *a* and *b* in Equation 4 are very close or proportional to μ and σ respectively, then the SNV pretreated result x_{SNV} comes out close to x_i.

MSC is more specifically derived to compensate the situation as described in Equation 4. This MSC algorithm (Isaksson and Næs, 1988; Esbensen, 2004) applies simple regression on the original spectral dataset and thus computes the vectors *a*' and *b*' at each wavenumber *i* (Equation 6).

Equation 6 $$y_i' = a' + b' x_i'$$

Therefore, if the physical interferences only cause the situation as described in Equation 4, then *a*' and *b*' essentially are *a* and *b*, and thus the MSC transformed spectra x_i' are perfectly reflecting x_i.

If the latent physical interference causes offset but not the multiplicative effect, changing the desired spectra to the presented spectra y_i: as in Equation 7

Equation 7 $$y_i = a + x_i$$

then 1st derivative can remove the term *a* by taking the difference between the intensity at two NIR wavelengths, resulting in a 1st derivative spectra d_i without any additive term (Equation 8).

Equation 8 $$d_i = (a + x_i) - (a + x_{i-1}) = x - x_{i-1}$$

If the latent physical interferences cause the offset no longer only a constant but also proportional to the wavelength (Equation 9),

Equation 9 $$y_i = x_i + bw_i + a$$

then the 2nd derivative can remove the terms *a* and *b* by taking the difference of the difference from 1st derivative, resulting in a 2nd derivative spectra (d_i ')'without any additive term (Equation 10).

Equation 10

$$(d_i^{'})^{'} = d_i^{'} - d_{i-1}^{'} = [(a + bw_{i+1} + x_{i+1}) - (a + bw_i + x_i)] - [(a + bw_i + x_i) - (a + bw_{i-1} + x_{i-1})]$$
$$= x_{i+1} - x_{i-1} + b(w_{i+1} - w_i) - b(w_i - w_{i-1}) = x_{i+1} - x_{i-1}$$

If the physical interferences cause the offset and the multiplicative coefficient associated with both x_i and wavenumber (Equation 11),

Equation 11 $y_i = a + bx_i + cw_i$ or $y_i = a + bx_i + cw_i + dw_i^2$

then EMSC develops a model with the assumption of the corresponding format and back-calculated x_i (Equation 12) (Martens and Stark, 1991; Martens et al., 2004; Esbensen, 2004).

Equation 12 $x_i = (y_i - a - cw_i)/b \quad x_i = (y_i - a - cw_i - dw_i^2)/b$

So far, with these theoretical deductions, almost every extraneous effect with its associated additive terms can be mathematically corrected and removed. The problem is to figure out what situation fits physical interferences in biomass compositional analysis case and accordingly make the modification.

Retrospecting the development of NIR technology, the original momentum came from the application rather than the theory. Following this approach, this study investigated and demonstrated the selection of data pretreatment methods from the experimental application results.

2.4.2 Multivariate analysis techniques

2.4.2.1 Principle component analysis (PCA)

PCA is utilized to decompose the original spectral dataset and re-locate the dataset in the newly developed orthogonal coordinates, principle components (PC) with maximum variance which reflects certain hidden phenomenon. Thus the capability of explaining the dataset variance is descending from the 1st PC to the latter PC (Kramer, 1998).

The only assumption in PCA (Esbensen, 2004) is that the original dataset

X can be split into a sum of a matrix product (PC), and a residual matrix E. The mathematical description is shown in Equation 13:

Equation 13 $$X = nPC + E = UL^T + E = u_1 l_1^T + u_2 l_2^T + ... + u_n l_n^T + E$$

where *U* is the score matrix which mathematically re-locates the data points in the newly composed co-ordinate system, and L^T is associated loading matrix which is the mathematical description of the new co-ordinates with respect to the original variables. Each data point has a set of scores by projecting onto the new coordinates, which are connected with the original variables via the loading vectors l_i. In Equation 13, each term $u_i l_i^T$ represents one principle component, and the number of PC is determined by the operator's judgment on the goodness-of-fit or how small he wants the residual to be (Gnandesikan, 1977; Osborne, 1986; Mark and Workman, 1987). But it is certain that the number of PCs is much less than the original variables, since significant collinearity exists in spectral data while the first PCs always explain larger variance.

2.4.2.2 Hierarchical clustering analysis (HCA)

The agglomerative hierarchical clustering algorithm is a method that classifies the data in hierarchy architecture via the degree of dissimilarity, and commonly displayed as a tree diagram called a dendrogram (Figure 3). It starts with every observation as one cluster, presented on the right side of the dendrogram. With certain computation algorithm, the observations with larger similarity fuse into one group, and this procedure continues until all the observations merged. Therefore, on the right end of dendrogram, the number of horizontal lines equals the number of the observations, while the left side only has only one line. The distance between two vertical split represents the dissimilarity of the corresponding two clusters. In this way, the internal structure and data grouping can be clearly observed. More detailed procedures are described as below:

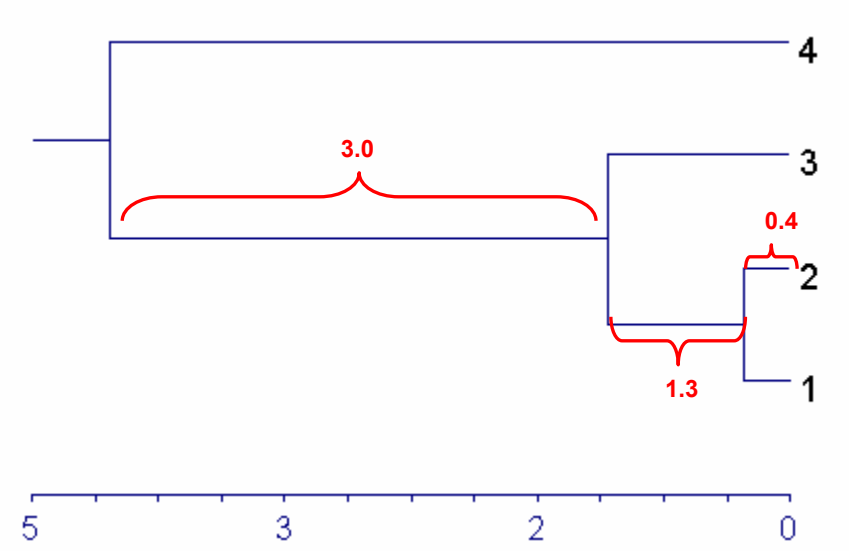


Figure 3 Illustration of dendrogram [1,2,3,4,5]

1. Initially four observations are four clusters on the right side of dendrogram.
2. 1 and 2 has the least dissimilarity, fusing first. The dissimilarity was calculated via the Euclidean distance (Appendix C), and equals 0.4. 1 and 2 merge as a new Cluster 1&2.
3. Calculate the distances between the Cluster 1&2 and other clusters, using Ward's minimum variance algorithm, and find Cluster 3 and Cluster 1&2 has the least dissimilarity (1.3). Therefore, Cluster 1&2 merges with Cluster 3, and a new Cluster 1&2&3 is formed.
4. Repeat step 3, calculating the distance between Cluster 4 and Cluster 1&2&3 (3.0).
5. The dissimilarity between 1 and 2 is 0.4; the dissimilarity between the Cluster 1&2 and 3 is 1.3, and the dissimilarity between the Cluster 1&2&3 and 4 is 3.0

2.4.2.3 Multiple linear regressions (MLR) with stepwise variable selection

MLR (De Noord, 1994) is a method for relating the variations in a response variable (Y-variable) to the variations of several predictors (X-variables), with explanatory or predictive purposes. If there are several response variables, MLR is performed multiple times so that each response variable has an associated model with the predictors. MLR requires the complete independence among the predictors, because multicollinearity causes the instability of the estimated regression coefficients. This requirement becomes the major drawback of MLR in NIR analysis, since multicollinearity is prevalent in NIR spectra. MLR is applicable when the spectrum contains only a small number of wavelengths/wavenumbers. The direct application of MLR usually violates the independence requirement of multiple regression; therefore, variable reduction such as stepwise selection should always be performed beforehand.

Stepwise variable selection uses the forward selection strategy modified with backward elimination (Johnson, 1998). It starts with no variable in the model, and adds the most significant variable to the model one each iteration, recalculates the significance of the updated candidate variables in the model and removes the non-significant ones from the model. This procedure iterates until no more variables can be entered or removed from the model at the predetermined significant levels, so program requires one significant level for adding the variable and another one for removing the variable. Usually larger significant level (α) is used for exclusion than inclusion to prevent an infinite loop (Steyerberg et al., 2000). Stepwise selection eliminates multicollinearity problem, and multiple linear regression consequently becomes valid to apply to the selected variables.

2.4.2.4 Principle component regression (PCR)

PCR is usually performed in a two-steps manner. First, PCA is conducted on the spectral dataset, resulting in a series of principle components (*PC*s) and

projecting the spectral dataset into an optimal coordinate system. Second, instead of the original variables, *PC*s are used to relate to the dependent variable dataset (chemical concentrations) with the least-square fitness in seek of the calibration coefficient set. It can be expressed as Equation 14 (Kramer, 1998):

Equation 14 $$C = K \cdot PC$$

Where: *C* is the chemicals dataset matrix; *PC* stands for pre-composed principle components matrix; and *K* is the regression coefficient matrix

2.4.2.5 Partial least squares regression (PLS)

PLS employs the same strategy as PCA not only on the spectral dataset (X-block) but also chemical side (Y-block) (Kramer, 1998). As a result, spectral data are expressed as projections onto the optimal spectral factors (coordinates), while the chemical data are projections onto a series of chemical factors (coordinates). Each pair of factors, composed of one spectral factor and its associated chemical factor is rotated or perturbed and adjusted to each other until minimum least square is achieved in the linear relationship between spectral dataset and chemical dataset. It can be expressed as:

Equation 15 $$C^{proj.}{}_i = K \cdot S^{proj.}{}_i$$

where $S^{proj.}$ is the projection of the spectral data onto the i^{th} spectral factor; C^{proj} is the projection of the associated chemical data onto the i^{th} chemical factor; and K_i is the coefficient between the i^{th} pair of spectral and chemical factor.

2.4.2.6 Leave-one-out cross-validation

Leave-one-out cross-validation is a complementary validation approach employed when the accessible sample set is not large enough to be separated into calibration set and validation set (Hildrum et al., 1996; Goodchild et al., 1998; Freitas et al., 2005; etc). First, one variable is on hold while the other variables are used in the model calibration. Then the excluded variable is used to validate the calibration model. The procedure continues until each of the samples has been kept out once, and the test statistics are the average of all the iterations.

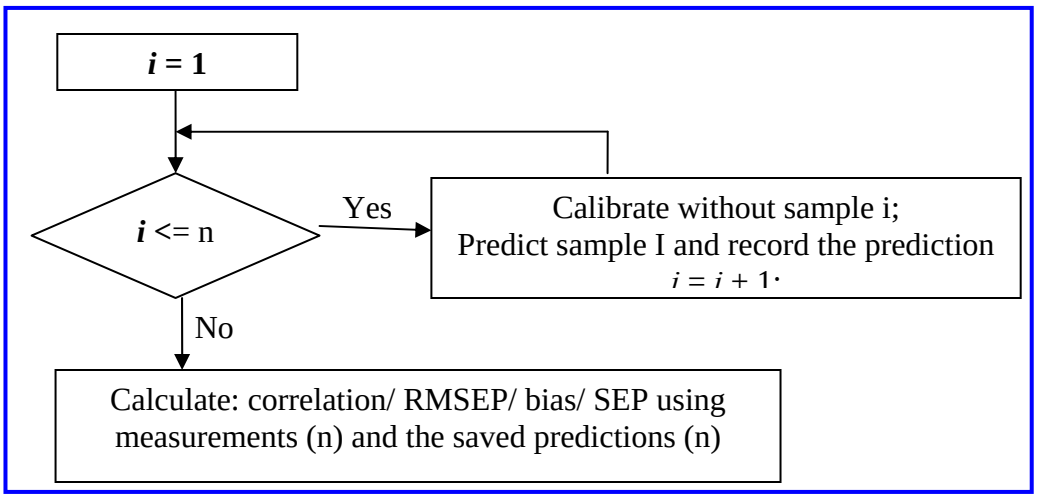


Figure 4 The flowchart of cross-validation, assuming the dataset includes *n* samples. The equations to calculate correlation, RMSEP, bias, SEP are listed I Appendix C.

Cross-validation mechanism can be more clearly presented in Figure 4.

Cross-validation is a conservative validation method, especially when applied to a small dataset with large variability. Since under this situation in every permutation the excluded sample is not involved in the calibration range developed on the rest samples, it is highly possible that the error is inflated.

2.5 Biomass properties

2.5.1 Chemical properties

2.5.1.1 Carbohydrates

Cellulose, the principle plant component, is a water-insoluble carbohydrate homo-polymer with large average molecular weight over 100,000 (McKendry, 2002). Its long linear chain repeats β-D-glucophyranosyl units joined by *β*-(1->4) glycosidic linkages (Figure 5), which can be also considered as a combined form of D-glucose.

Unlike linear and stereoregular cellulose structure composed of the same unit, the hetero-polymer hemicellulose is a highly branched chain composed of

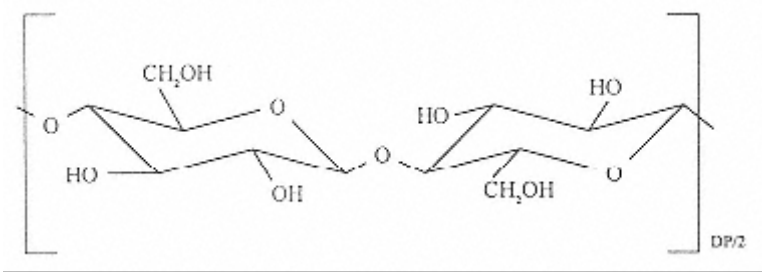


Figure 5 Cellulose repeating β-1,4-linked anhydrocellobiose unit (Fennema, 1996)

different monomer sugar units. Hemicellulose has a lower average molecular weight (<30, 000) (McKendry, 2002), and is amorphous and easily hydrolyzed into D-xylose, L-arabinose, D-glucose, D-galactose and D-mannose.

Although cellulose and hemicellulose are the existing carbohydrate format in the biomass plant, the analytes associated with carbohydrates throughout this study are glucose, xylose, galactose, arabinose, and mannose.

Several reasons account for using these monosaccharides instead of cellulose and hemicellulose. First, they are the downstream products and directly measured in the standard wet chemistry analysis. Second, monosaccharides, especially glucose, are the direct reactants for the ethanol fermentation, hence the knowledge of their values is more straightforward and instructive for industrial use. Also, these five monosaccharides are the hydrolysis products of cellulose and hemicellulose, so it is possible to calculate cellulose and hemicellulose contents based on them.

Among these five simple sugars, xylose and arabinose are five-carbon sugars while glucose, galactose, and mannose are six-carbon sugars. Their open-chain molecular formulas are presented in Figure 6.

2.5.1.2 Lignin

Lignin is amorphous aromatic macro polymer with very complex structure (Appendix D). It functions to bind the individual cell together and harden the cell tissue so that to concrete the plant structure (Mckendry, 2002). The phenyl-

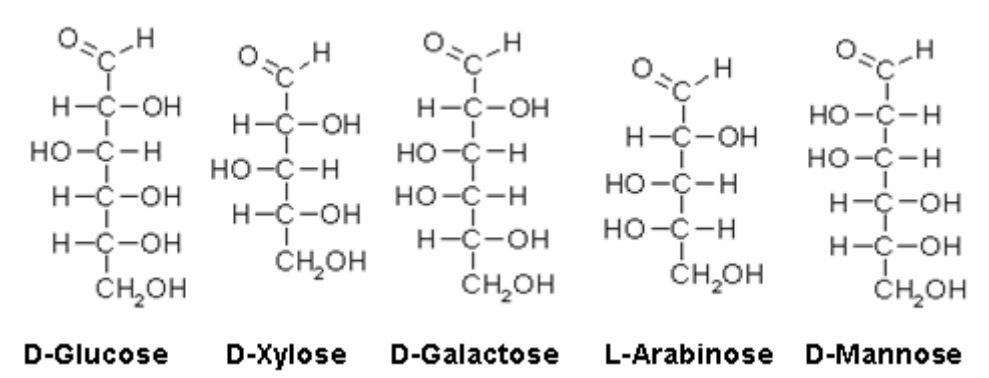


Figure 6 Fisher open-chain molecular formulas of five simple sugar analytes commonly encountered in biomass

propane units are connected via varieties of linkages (Figure 7).

Therefore, both carbohydrates and lignin are organic compounds, consisting of varieties of chemical bonds (C-C, C-O, C-H, O-H, C=C, C=O). These chemical bonds absorb NIR radiation, presenting the accumulative results at each wavenumber cross the entire NIR region in spectra.

2.5.1.3 Ash

Ash is mainly composed of the inorganic compounds, such as SO_2, CaO, K_2O and Al_2O_3 and some trace minerals, such as Cr, Ni, Co, etc. (Demirbas, 2005). There is no NIR absorption occurring to most of the inorganic matters in ash (Lestander and Rhén, 2005). However, it is possible to use NIR technique to probe ash content (de Aldana et al., 1996; Lestander and Rhen, 2005; Znidarsic, et al., 2005, Cozzolino et al., 2006), since its existence may correlate with other constituents thus indirectly leading to the NIR spectral variation.

2.5.2 Heating values

Heating value provides direct guidance for the bio-power plant in estimating the energy potential that can be transferred from certain feedstock or fuel products. Higher heating value (HHV), or gross calorific value, is the potential combustion energy when water vapor from combustion is condensed, thus taking into

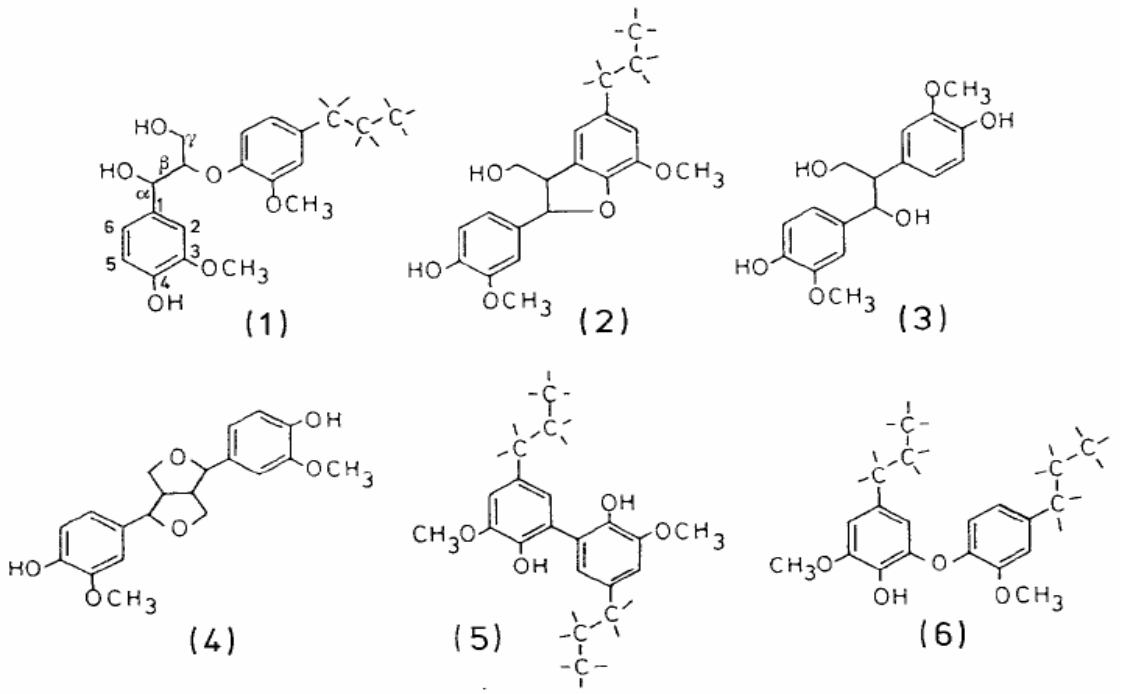


Figure 7 The principal linkage modes between the phenylpropane units in lignin macromolecules. (1) guaiacylglycero-β-aryl ether substructure (40-60%), (2) phenylcoumaran (10%), (3) diarylpropane (5-10%), (4) pinoresinol (<5%), (5) biphenyl (5-10%), (6) diphenyl ether (5%) (Higuchi, 1990).

account the latent heat of vaporization of water in the combustion products. HHV is obtained when all the combustion products are cooled down to the original temperature (usually 25 ºC).

HHV measurements basically are based on oxidization reaction, and all the reactants are the chemical constituents in biomass, and thus HHV essentially is associated with the chemical bonds characteristics and quantities. Therefore, it is reasonable and possible to use NIR spectra to predict HHV.

2.6 Objectives

This study included both chemical composition investigation and FT-NIR modeling in terms of biomass chemical composition and HHV as well. The objectives were:

- to analyze the chemical composition of corn stover and switchgrass, in terms of glucose, xylose, galactose, arabinose, mannose, lignin, and ash, and investigate the variability existing in botanic fractions and cultivars for directing future planting and industrial feedstock selection;
- to determine the best modeling configurations for biomass NIR study in three aspects: 1) the selection between two spectral transform algorithms for better linearity between chemical composition and NIR spectra; 2) the selection of the optimal spectral pretreatment method among the nine most frequently used methods in previous NIR studies; 3) the selection of the optimal multivariate regression method among MLR, PCR and PLS;
- to develop the individual FT-NIR-based models for corn stover and switchgrass to rapidly and accurately predict their chemical composition respectively;
- to test the hypothesis that a general model is valid for both corn stover and switchgrass in probing their chemical compositions, and to develop the general FT-NIR predictive model for both species if the hypothesis holds;
- to test whether the general model has the potential predictive capability to predict the chemical composition of other species;

- to investigate the potential of using FT-NIR techniques to predict HHV of the biomass feedstock.

CHAPTER III LITERATURE REVIEW

3.1 NIR Applications in related disciplines

There are a great many applications of NIR in predicting chemical properties in many agricultural disciplines with different focuses.

In the animal feed and forage areas, studies concentrate more on the nutrition content and thus many indexes such as protein, fibers, and starch are of major interest. Lucerne forage (Martens, et al., 1984; Flores Pires, et al., 1998), legumes (Martens, et al., 1984), maize stover (Albanell et al.,1995; Cozzolino et al., 2000; Wei et al., 2005), maize grain (Melchhinger, et al., 1990), mixed pasture, and whole plants have been investigated for acid detergent fiber (ADF), neutral detergent fiber (NDF), and protein, which has been studied with some other species, including stargrass/ bermuda grass (Brown, et al., 1990), and temperate grass silage (Smith and Flinn, 1991). Although these materials were all agricultural crops or herbaceous plants that could also be used as bio-industry feedstocks, the predictive models developed by these studies are not appropriate as a guide for bio-fuel production. Protein, ADF and NDF are no longer of major interest in bio-energy industry, and instead, monosaccharides, lignin and even the small amount of ash have been focused on, because these constituents are directly related with bio-fuel production, or bio-energy production. Lignin has been the anaylte in the previous NIR studied in the forage discipline, and the reported reference contents were 6.1% for legumes (Marten et al., 1984), 3.3% for maize stover (Zimmer, et al., 1990), 1.9-8.3% for mixed pasture species (Garcia-Cuidad et al., 1993) respectively. However, these values were significantly different from the commonly detected range 10-25% in bio-renewable energy area. The reason is that the interested lignin in forage industry is actually acid detergent lignin, which is a part of total lignin content (acid soluble lignin and acid insoluble lignin) Also, the wet chemistry method for acid detergent lignin (Donnelly and Wear, 1972; Edwards, 1973; Jung, 1989) is different from the one commonly employed in forestry and bio-energy discipline (ASTM E-

1721-95, 1995, Sluiter et al., 2006). Therefore, although some agricultural residuals and herbaceous grasses were studied before, since the analytes differ and reference methods differ, those study results are not applicable to current biomass energy programs, and new predictive model development is demanded.

Woods have similar composition to biomass with the major constituents: cellulose, hemicellulose, and the non-destructive NIR research has been conducted for years in forestry area, which is very instructive to renewable energy feedstock biomass. Lignin is a major concern for wood industry and thus has been focused on. Balleirini dos Santos et al. (2006) reported their prediction of lignin content with a 93% correlation and 0.55% error rate via NIR on Eucalyptus. Poke and Raymond (2006) predicted chemical composition of solid woods using an existing calibration NIR model of ground woods; however, the prediction was really poor. Therefore, they specifically developed a calibration model for solid wood and achieved the R^2 from 0.67 to 0.87 for total lignin, cellulose and extractives yet the acid-soluble lignin was not well predicted. Second derivative NIR spectral data at 1673 nm was correlated to lignin content and a correlation of 0.956 for the milled spruce wood samples and 0.984 for the fungi-treated samples were achieved (Schwanninger et al., 2004). Many other chemical properties, including carbohydrates, have also been studied for many wood species (Michel, 1988; Michell and Schimlec, 1994; Schimleck et al., 2000; Kelley, 2004; Lestander and Rhen, 2005).

All the studies above were conducted using dispersive NIR spectrophotometers. With the development of powerful microcomputer Fourier transform function was applied to NIR instrumentation, which enhanced the capability and improves the accuracy of NIR techniques. An FT-NIR spectrophotometer utilizes an interferometer to modulate the signal, and each wavelength has a distinctive modulation frequency, which is to be transformed immediately via the computer into its actual electromagnetic frequency. Given all the features overcoming the dispersive NIR (Bell, 1972; Griffiths and de Haseth, 1986; Burns, 1997), FT-NIR has the potential to finally dominate NIR instrument

market. Peirs et al. (2002) did a parallel comparison between FT-NIR and dispersive NIR spectroscopy on investigating apple quality, where they claimed that FT-NIR overcomes dispersive NIR not only from theoretical and mechanical reasons but also from the experiment result. Due to its late start, the applications of FT-NIR in biological materials are insufficient. Much less efforts have been made using FT-NIR, compared to plenty of studies using dispersive NIR. Applying S-Golay second-derivative (Savitzky and Golay, 1964) to the raw FT-NIR spectra, Schwanninger et al. (2004) successfully differentiated the thermally treated wood. Manley et al. (2002) announced that scanning instruments (FT-NIR) overcomes the weakness of the filter instrument (dispersive NIR), so they chose FT-NIR to study whole wheat flour and successfully determined the kernel hardness, protein and moisture content. The feasibility of applying FT-NIR in determining extractive and phenolic contents in hardwood of larch trees was investigated by Gierliner et al. (2002), and the results proved that FT-NIR was a reliable and accurate approach for wood extractive determination.

The successful applications of compositional analyses using NIR techniques in the related disciplines shows promising future of applying the techniques to the investigation of renewable energy feedstock biomass. However, it is necessary to do the research for biomass specially. First, different industries use different standards and criteria to investigate the feedstock, and the focuses vary even as for the same material. For instance, when the corn stover is used as the feedstock in forage industry, fiber and protein will be of interest, while in bio-renewable energy industry, sugars and lignin are the major concerns. Second, the physical and chemical characteristics of feedstock vary by the material used in different industries. For examples, eucalypts have larger portion of lignin than grasses; cereal grains have great starch content while the switchgrass dose not. Therefore, previous NIR studies in other areas show the feasibility yet do not ensure the conclusion of NIR application on biomass. Furthermore, most of the previous NIR studies were developed on dispersive NIR spectrometers, including some studies on corn stover and stover-derived feedstock (Hames et al., 2003).

They are not compatible to FT-NIR spectrometers, which have higher resolution. So far, few NIR applications on switchgrass have been reported, and neither corn stover nor switchgrass has been studied using FT-NIR. Therefore, it is worthwhile to verify the feasibility of FT-NIR analysis on corn stover, and further extent the analyses to switchgrass composition prediction.

3.2 Discussion on calibration dataset

An accurate and robust model requires a diverse yet typical calibration dataset, covering as large variability as possible yet still representative enough for the investigated subject. A lot of studies achieve the large variation requirement through collecting samples from different locations or different time frames; however, sometimes, these approaches are not very easily accessible. Lestander and Rhen (2005) used NIR spectroscopy and bi-orthogonal PLS regression to model moisture and ash content as well as gross calorific value in ground samples of stem and branches wood of Norway spruce. The study presented the big variation between stems branches, and suggested different botanic parts of plant vary a lot in the chemical composition. Crofcheck and Montross (2004) studied enzymatic hydrolysis on three fractions of corn stover, cobs, leaves, and husks, with the attempt to use fractions with higher glucose potential. The result of this study shows that there are differences among the fractions, and infers that manual separation could be a new pathway to introduce larger variation to NIR calibration.

Now that the variability has been emphasized to enhance model robustness, a question will be naturally associated: can the variation be created via introducing a quite different material? For instance, in predicting lipids large variation can certainly and easily be achieved by combining oats and potato chips in calibration dataset, but the significance and representation of the result is questionable and meaningless scientifically and statistically. This is an extreme case, but what if using two species that are close in the chemical composition?

Throughout literature, most NIR studies was conducted on one species, for example, as for animal feeds, lucernes (Martens, et al., 1984), legumes

(Smith and Flinn, 1991) or maize (Melchinger, et al) were investigated separately through different studies; similarly, as for woods, pulpwood (Schimleck et al., 2000), eucalyptus (Michell and Schimlec, 1994) and Norway spruce (Lestander and Rhen, 2005) were separately studied. These suggest that most researches hold the idea that different models should be developed respectively for different species, which was stated clearly by Hames et al. (2003). While most studies developed each method specifically for one species, there were several studies that developed models for a range of species. Xiccato et al. (1999) used NIR reflectance spectroscopy to predict chemical constituents, digestibility and energy value of many feeds for rabbit, including alfalfa meal, dried beet pulp, sunflower meal, wheat bran, whole soy bean, grains (barley, wheat) and wheat straw. The prediction results were not very satisfactory, with R^2 ranging from 0.25 for organic matter to highest value 0.93 for ether extract. This result was discussed to be probably due to too many species in calibration dataset, which again indicate that each species probably should be studied individually in the first place. Other studies (Brown, et al., 1990; Carcia-Cuidada, et al., 1993) also showed the prediction results from many species were not as good as those from individual species. However, Sanderson et al (1996) used a calibration dataset, which was composed of varieties of woody and herbaceous species, and achieved good prediction result of ash, lignin, arabinose, xylose and N from the calibration dataset.

This controversy introduced an objective to this FT-NIR study. Is one general model valid and appropriate for predicting both biomass species (corn stover and switchgrass)? Therefore, other than developing individual models for corn stover and switchgrass individually, another objective of this study is to justify and verify one hypothesis: one general model can be developed on the both species and thus be used to predict either corn stover or switchgrass chemical composition.

3.3 Data processing

3.3.1 Data pretreatment methods

When NIR techniques are in use to access chemical composition information, the ideal scenario occurs when the differences among spectra are exclusively attributed to the chemical properties. However, many physical phenomena exist, especially in diffuse reflectance sampling, interfering with NIR spectral acquisition of chemical information. Although there are many pretreatment methods for NIR spectra to remove the physical interferences, no consensus has been established in the literature. The most frequently used methods are 1st derivative (eg. Flinn et al., 1996; Church et al., 1999; Nousiainen et al., 2004; Confalonieri et al., 2004; Veraverbeke et al.,2005; Liu and Ying, 2005; Dou et al., 2006), 2nd derivative (eg. Sverzut et al., 1987; Hong et al., 1996; Liu et al., 1998; Rodriguez-Saona et al., 2001; Tosi et al., 2003;), SNV (eg. Sánchez et al., 2004; Esbensen, 2004), MSC (eg. Shimoyama et al., 1999; Czarnik-Matusewicz et al., 1999; Baianu et al.,2004; Sato et al., 2003; Munck, 2006), and EMSC (eg. Schonkopf et al., 1992; Martens and Stark, 1991; Saiz-Abajo et al., 2005). Also, the combination usage of SNV with either 1st or 2nd derivative have been theoretically addressed (Fearn, 2000). There is obviously not a universal pretreatment method that we can take it for granted in the biomass FT-NIR research. Therefore, a search for pretreatment method in biomass application is necessary.

3.3.2 Spectral ordinate: *K-M* or *log(1/R)*?

Many studies used *log(1/R)* as the transform equation (Martens, et al., 1984; Liu, 1996; Walsh, et al., 2000, etc). Especially, Burns (1997) claimed that *log(1/R)* gives better linearity association to concentration, and thus is more useful when matrix absorbs at the same wavelength as the analytes. It was later argued that *K-M* definitely outweighs *log(1/R)* (Dahm et al., 1995). With the same idea as Dahm et al., many studies (Fardim et al., 2002; Andres, 2005; Fardim et al., 2005; Morgano et al., 2007; etc.) used the spectra with the *K-M* transform

algorithm in their quantitative analyses.

Given the fact that NIR calibration does not precisely obey Beer's Law and some multivariate methods are able to reconstruct the data for the goodness of fit, it is difficult to conclude, for biomass materials, whether *K-M* or *log(1/R)* provides better linear relationship between spectral dataset and the chemical dataset. This issue will be examined in this study via pair-wise comparisons.

3.3.3 Multivariate analyses

PCA is the most commonly used multivariate method in qualitative analyses of NIR spectra (Esbensen, 2004). PCA is used to decompose the series of spectral data, and recomposed with much fewer numbers of principle components that represent the maximum of the data variance, coherently realizing the data reduction. Compared to the wide utilities of PCA, HCA, which can also be used as NIR qualitative tool, has much less applications in NIR research; especially, applying HCA to the spectral pretreatment investigation has rarely been seen in the literature.

As for the quantitative NIR analyses, a great many studies (Albanell et al., 1995; Cozzolino et al., 2000; Hames et al., 2003;Schimleck et al., 2000; Lestander and Rhen, 2005; Wei et al., 2005; etc) used PLS to develop the regression model. Other than this most frequently used method, some other approaches also have been utilized in the quantitative NIR analysis, such as PCR (Isaksson at al, 1995; Sun, 1996; Chang et al., 2001; Via et al., 2003; Medendorp et al., 2006; etc) and MLR (Otsuka et al., 2000; Suehara, 2004; Ito, 2007; Szlyk et al., 2007; etc). Facing these multivariate methods especially the regression options, this study has objective to determining the best functional regression methodology for the FT-NIR biomass chemical composition modeling.

CHAPTER IV MATERIALS AND METHODS

This chapter provides the detailed information associated with the biomass materials used in this study (Section 4.1), the experimental procedures and methods (Section 4.2) and the experiment designs to achieve the multiple objectives (Section 4.3).

4.1 Biomass materials

4.1.1 Corn stover

Corn stover samples of the cultivar DeKalb DK64-10RR were collected from TN Knoxville Agriculture Experimental Station in May 2006, and air dried in the lab. Different botanical parts have different structure, serving different biological functions, and thus possibly dissimilar ratio of chemicals. Therefore, botanic fractions were utilized in this study to create the variability for the calibration model accuracy and robustness. The preliminary results showed the feasibility of using botanic fractions of corn stover to create the variability. The whole corn stalk was studied, together with the manually separated botanical parts, namely nodes, piths, rinds, sheaths, leaves and husks (Figure 8).

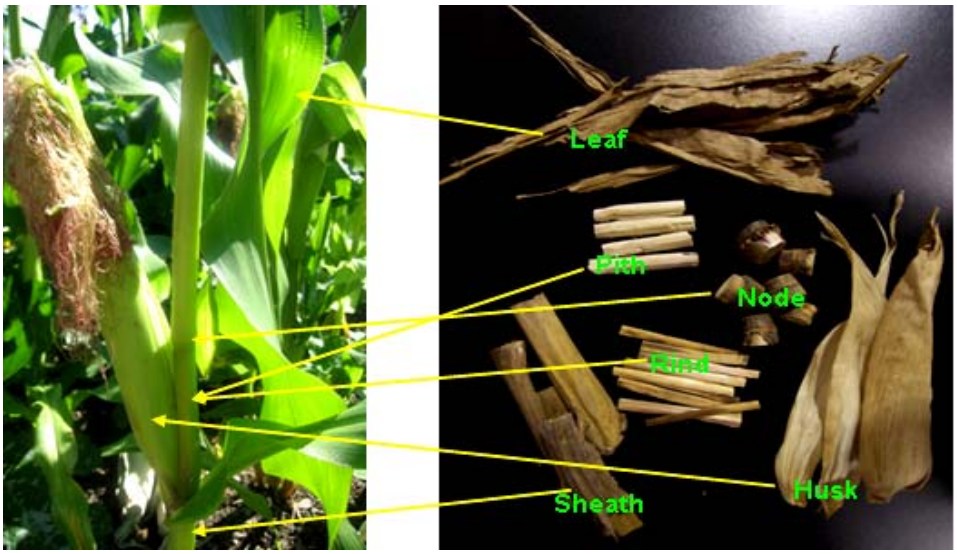


Figure 8 Manually separated botanical fractions of corn stover

This alternative solution breaks through geographical and temporal limitations and even endows researchers the flexibility to prepare certain samples with desired compositional proportion. The meaning of separation lies not only in manually creating large variability, but also in making better use of corn residuals. For example, botanic parts with higher glucose content should be selected as feedstock for bio-ethanol production, while those with lower sugar content are better to be left in the field to provide sufficient erosion control. In this way, ethanol production cost is reduced and the yield is increased.

In addition, stover samples of a sweet corn cultivar (Incredable) were collected at the same time in small quantities and only used in the pretreatment experiment (refer to Section 4.3.2.2.1)

4.1.2 Switchgrass

There were six switchgrass cultivars available at TN Agriculture Experimental Station, including Cave-in-Rock, Alamo, Kanlow, Shelter, NC1-16, and NC2-16. They were collected in August, 2006, and stored in the lab. The morphological heterogeneity of the nodes, internodes, and leaves (Figure 9) suggests that chemical composition distinction would exist among these switchgrass botanic parts. The aerodynamic partition of these botanic parts was proved to be feasible (Klasek, 2006). However, further chemical composition investigation with respect to the botanic fractions is rarely seen in literature. Therefore, it was worthwhile to investigate these major botanic fractions, which meanwhile provide more variability for the model calibration. Alamo and Kanlow were further separated into botanic parts, because a previous study (Lemus et al., 2002) reported that these two varieties produced the highest biomass yield among twenty switchgrass populations, with lowest ash contents. Consequently, these two cultivars have greater potential to be predominantly grown and thus it was worthwhile to further study them.

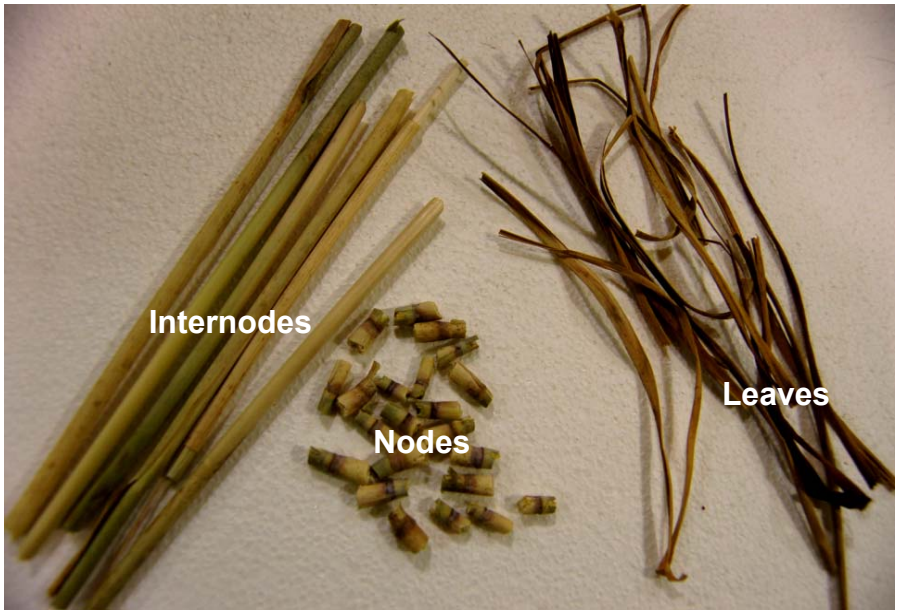


Figure 9 Manually separated botanical fractions of switchgrass

4.1.3 Wheat straw

The wheat straw samples were harvested from UT dairy site in Blount County, TN, in 2005 summer. Wheat straw was incorporated in model validation to test whether the hypothesized general model was capable to predict the chemical composition of a third biomass species. Several reasons accounted for the selection of wheat straw as part of the independent validation data: 1) according to Department of Energy database (http://www1.eere.energy.gov/biomass/feedstock_databases.html), wheat straw chemical composition is very close to that of corn stover; 2) it was easily accessible; 3) a quick check of the associated spectra showed that all of them located within the calibration spectral dataset.

If the prediction of wheat straw was validated, it means that the developed model could be extended to cover more biomass species as long as their chemical compositions are within the calibration range. And it would suggest as well that the physical characteristics and genotype differences can be corrected by spectral pretreatment as well.

4.2 Methods

4.2.1 Sample preparation

Biomass sample were stored in the laboratory at the ambient temperature (20-30 °C) and less than 50% relative humidity for at least four months. All the samples were air-dried to moisture content within 10%. Wiley Mini Mill (Thomas Scientific) was utilized to grind the biomass samples. Three sieve sizes (20 mesh (0.85 mm opening), 40 mesh (0.425 mm opening) and 60 mesh (0.25 mm opening)) were used for the pretreatment selection experiment (refer to Section 4.3.2.2.1). Other than that experiment, all the other samples used in this study were prepared through 40 mesh sieve. Approximately 1 g ground biomass were sampled in a non-absorbing glass vial (PIKE Technologies, Madlson WI), and dried in a 105°C convection oven until a constant weight was achieved.

4.2.2 Date collection

4.2.2.1 Overview of composition analysis model calibration

Oven-dried samples were cooled to room temperature in a desiccator before the FT-NIR spectra were acquired. Immediately after the spectra acquisition, the sample was subjected to wet chemistry analysis and the chemical data on glucose, xylose, galactose, arabinose, mannose, lignin, and ash were collected. The wet chemistry analysis and FT-NIR sampling procedures are described in detail in Section 4.2.2.3. During the transportation between FT-NIR sampling and wet chemistry analysis, the samples were kept in the vial and well sealed to keep out the moisture. A complete set of one sample was composed of both chemical and spectral data. These procedures (Figure 10) were repeated for all the samples; and finally, the models were developed based on the complete dataset using the selected multivariate analysis method. These are the general experimental procedures for the development of FT-NIR composition analysis model.

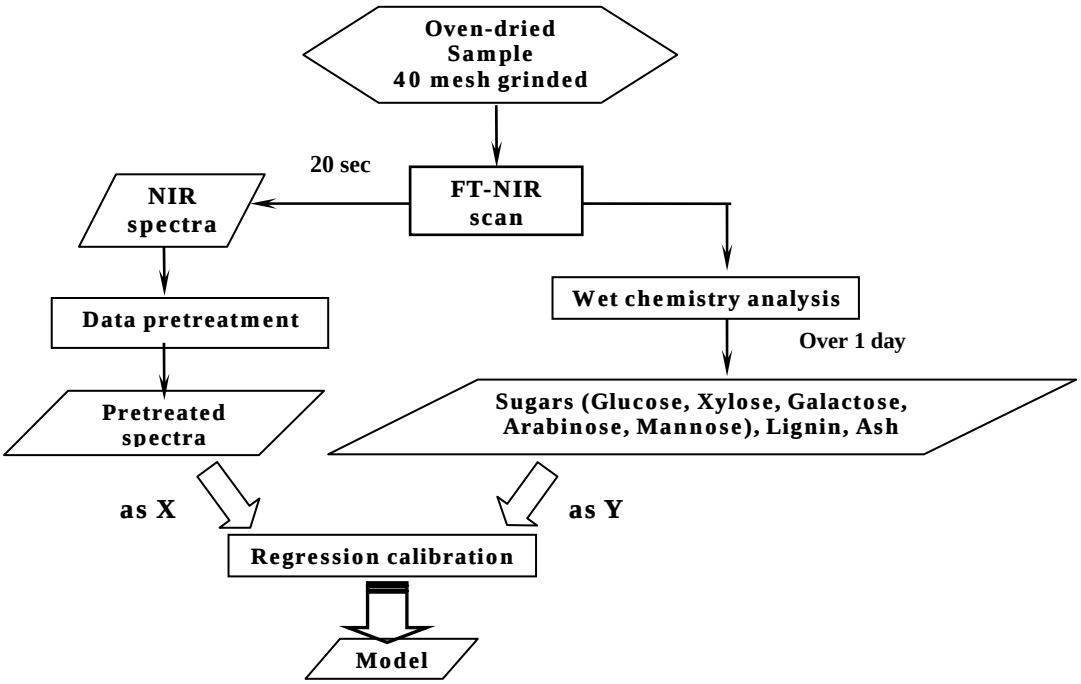


Figure 10 Scheme of data collection and model development

4.2.2.2 FT-NIR sampling

FT-IR spectrometer Excalibur 3100 (Varian Inc., Palo Alto, CA) is equipped with NIR IntegratIR integrating sphere accessory (PIKE Technologies, Madlson WI) and a build-in high-speed, low-noise, indium-gallium-arsenide (InGaAs) detector. Background spectrum was collected each day using a diffuse gold reference plate and automatically included in the calculation of each sample spectrum in order to minimize the atmosphere variation effect. Two algorithms (*log(1/R)* and *K-M*) were used to transform the diffuse reflectance radiation to a absorbance unit and thus the spectra were recorded with the ordinate of *log(1/R)* and *K-M*. The sample vial was mounted on the FT-NIR sample holder. NIR light beam shot from underneath to the samples through an optimized borosilicate window with 10 mm diameter (Figure 11). All FT-NIR diffuse reflectance spectra were collected at the resolution of 8 cm^{-1} over the 4000-10000cm^{-1} spectral region and one spectrum was formed by averaging sixty-four scans.

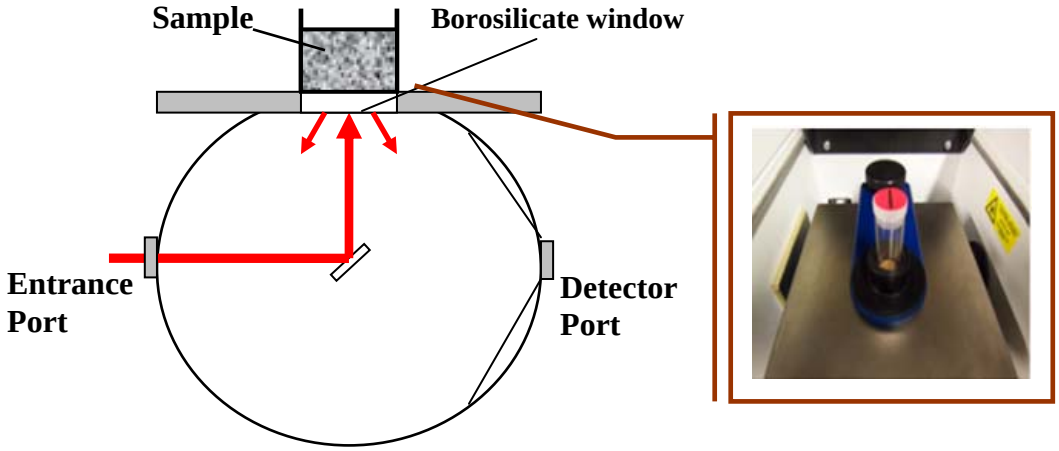


Figure 11 Optical geometry of an integrating sphere for NIR diffuse reflectance sampling

Three spectral samplings were conducted for each vial of sample with vigorous shaking between samplings so that the diffuse reflectance sampling represents more portions of the biomass powder within the vial. The average spectrum was calculated out of these three spectral samplings in the Unscrambler 9.2 (statistical software), and coherently used in the multivariate analysis. In this way, the spectrum was more representative to the whole population within the vial.

4.2.2.3 Wet chemistry analysis

Immediately after FT-NIR spectral acquisition, the sample was subjected to wet chemistry analysis (Figure 12), basically following NREL Laboratory Analytical Procedure "Determination of structural carbohydrates and lignin in biomass" for hydrolyzed monosaccharides and lignin measurement, and ASTM "Standard Method for the Determination of Ash in Biomass" for ash measurement (ASTM E1755-01). The procedure is briefly described as below:

1 Dry biomass (300.0±10.0 mg) was weighed into a crucible and subjected to 575 ºC furnace for 12 hr to obtain ash content.

2 Another 300.0±10.0 mg biomass were weighed out into a pressure tube, and then well mixed with 3.00±0.1 ml 72%w/w sulfuric acid. The tube then stayed

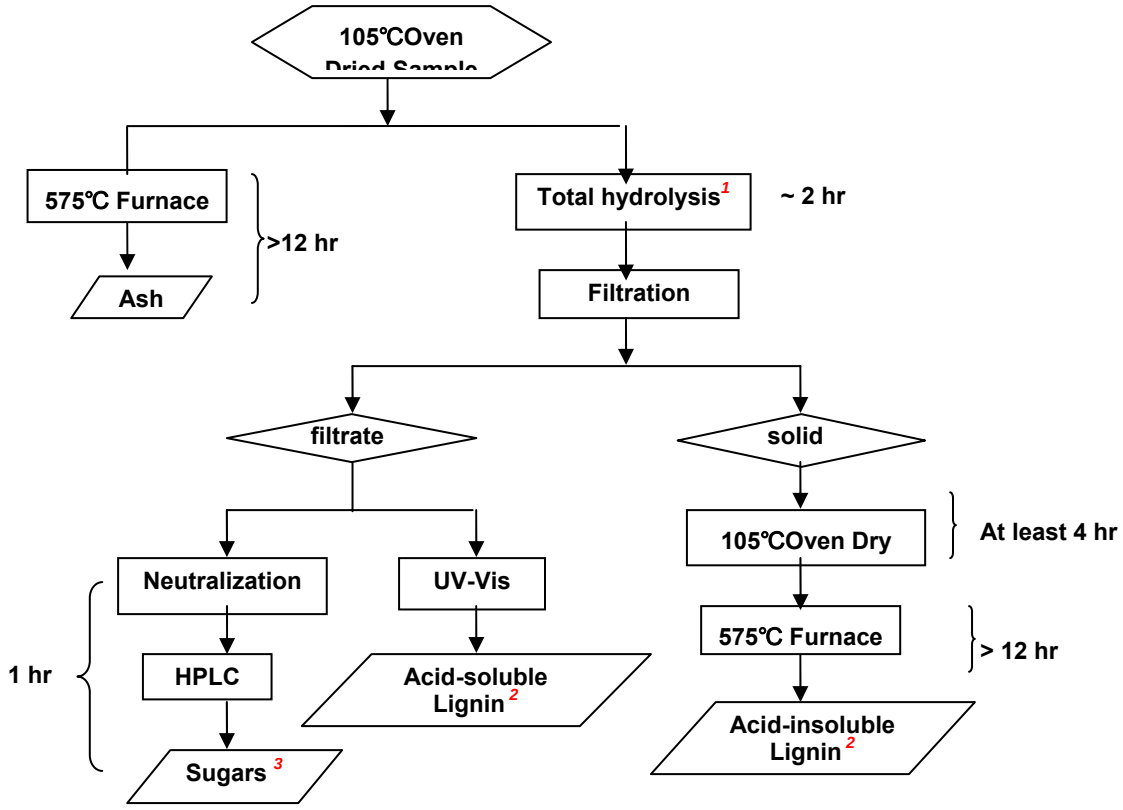


Figure 12 Wet chemistry analysis process of ground biomass samples

1. Include 2 steps: 1st concentrated acid hydrolysis and 2nd dilute acid hydrolysis.
2. Total lignin content is equal to acid-soluble lignin plus acid-insoluble lignin.
3. Detectable sugars include glucose, xylose, galactose, arabinose, mannose.

in 30 ±3 ºC water bath for 1 hr; during this time span, samples should be stirred at least every 10 min to make reaction thorough.

After the completion of this concentrated acid hydrolysis, the tube was removed from the water bath and 84.00±0.04 ml deionized water was added to dilute the acid to 4%w/w concentration. The sealed tube was placed in Napco 8000-DST bench top autoclave (Winchester, VA) at 121 ºC for 1 hour. During this dilute acid hydrolysis step, a set of sugar recovery standards (SRS) was conducted to correct for the possible loss due to destruction of sugars during dilute acid hydrolysis step. D-(+)glucose, D-(+)xylose, D-(+)galactose, L(+)arabinose, and D-(+)mannose (purchased from Sigma-Aldrich, St. Louis, MO) that closely resembled the real situation in biomass samples were weighed out to a pressure tube, and added with 10.0 mL deionized water and then 348.5 µL of 72% w/w sulfuric acid. The tube was placed in the autoclave together with the sample tubes.

3 The autoclaved hydrolysis solution was vacuum filtered through a previously weighed filtering crucible (Coors porcelain, medium porosity). Crucible with filtrated solid were 105 ºC oven-dried for at least 4 hours before being sent to 575 ºC furnace for another 12 hours. Weight difference (before and after furnace) was computed as acid-insoluble lignin.

4 A portion of the filtrate was sent to Schimatzu UV-1700 UV-spectrophotometer for acid-soluble lignin detection, and UV-absorbance was acquired at wavelength 320 nm and used for the calculation of acid-soluble lignin content. Afterwards, total lignin content was computed as the sum of both acid-soluble and acid-insoluble lignin.

5 A portion of filtrate was neutralized using calcium carbonate (ACS reagent grade) until pH reached 5-6. The neutralized filtrate was first vacuumed filtered and then further filtered through 0.2 µm syringe filter before being injected into HPLC. BioRad Aminex HPX-87P column pre-protected by ionic-form H^+/CO_3 deashing guard column was used to separate the monosaccharides before

entering the Waters 410 refractive index detector. Column operating condition was set at 80 ºC with the flow rate at 0.6 ml/min. The real-time signal outputs were monitored, and all the integrations were performed using the software GC Chemstation (Agilent Technologies, 1990-2001). These HPLC procedures were used for the measurement of the five monosaccharides: glucose, xylose, galactose, arabinose, and mannose.

4.2.2.4 Data processing and multivariate analyses

The Unscrambler 9.2 was used for all the spectral data pretreatments, PCA, PCR, PLS, and MLR in this study. Some other multivariate analyses, such as HCA and stepwise variable selection were conducted in the statistics software NCSS 2004. The use of multivariate methods is dictated specifically in the following section.

4.3 Experiment designs

4.3.1 Water bands study and justification for dry sample usage

The grinded switchgrass sample (1.20±0.05 g) was measured into a FT-NIR sampling vial and oven-dried until a constant weight was reached. FT-NIR spectrum was collected on this vial of moisture-free sample. Then it underwent isotherm sorption process at 25°C and 100% relative humidity, which was realized by being placed in the headspace of air-tight container with water at the bottom. After six hours, it was taken out of the container, acquired another FT-NIR spectrum and put back in. The next FT-NIR spectral sampling was the next day, and repeated for another two acquisitions. Afterwards, the time interval was extended to two days for four more acquisitions. All the weights were recorded, and the associated moisture contents were thus calculated. Nine observations in total were collected.

The scope of this experiment was to investigate how moisture content affected the NIR spectrum, and thus to determine the FT-NIR sampling condition for the modeling in this study.

4.3.2 FT-NIR predictive model development

4.3.2.1 Calibration dataset design

To develop the switchgrass FT-NIR model, three replicates were collected for each of six cultivars, Cave-in-rock, Alamo, Kanlow, Shelter, NC1-16, and NC2-16 respectively (Table 2). Three replicates for each of the three botanic fractions (leaves, nodes, and internodes) were performed for Alamo and Kanlow, respectively. Therefore, in total, thirty-six observations (Table 3) were collected, including both spectral and chemical information.

To develop the corn stove FT-NIR model, DeKalb DK64-10RR was used and manually separated into six botanic parts: husks, piths, rinds, nodes, sheaths and leaves. Five replicates were collected for each of these six botanic parts and for the whole stalk as well (Table 4). Therefore, totally, thirty-five observations were obtained.

All the 36 switchgrass samples and 35 corn stover samples were combined to compose the calibration dataset (71 samples in total) for the general model hypothesis testing, and a general model was justified and then developed.

Table 2 Experiment design of switchgrass data collection: 3 repetitions for each of the 6 switchgrass cultivars, and 3 repetitions for the three botanic parts of Alamo and Kanlow, respectively.

	Cave-in-rock	*Alamo*	*Kanlow*	*Shelter*	*NC1-16*	*NC2-16*
Whole Stalk	3	3	3	3	3	3
Leaf	-	3	3	-	-	-
Node	-	3	3	-	-	-
Internode	-	3	3	-	-	-

Table 3 Experiment design of corn stover data collection, 5 replicates for each of the six botanic fractions of corn stover and also 5 replicates for the whole stalk

Whole Stalk	*Husk*	*Pith*	*Rind*	*Node*	*Sheath*	*Leaf*
5	5	5	5	5	5	5

Table 4 Data denotation of the experimental design for pretreatment selection [1,2,3]

	20 mesh	*40 mesh*	*60 mesh*
Husk (DK64-10RR)	Hr20-1	Hr40-1	Hr60-1
	Hr20-2	Hr40-2	Hr60-2
	Hr20-3	Hr40-3	Hr60-3
Husk (Incredable)	Hs20-1	Hs40-1	Hs60-1
	Hs20-2	Hs40-2	Hs60-2
	Hs20-3	Hs40-3	Hs60-3
Switchgrass (Shelter)	G20-1	G40-1	G60-1
	G20-2	G40-2	G60-2
	G20-3	G40-3	G60-3

1. Three biomass categories (ROWs) are marked as Hu, Hs, and G
2. Three sieve sizes are shown as COLUMNs and are marked as 20, 40, 60.
3. Three replicates were conducted for each biomass variety, following 20, 40, 60 mesh grinding subsequently.

4.3.2.2 Discussion on the best modeling configuration

4.3.2.2.1 Investigation of spectral pretreatments

Three replicates of the husks of two corn cultivars (Incredable and DeKalb DK64-10RR) and the internodes of one switchgrass cultivar (Shelter) were subjected to gradient grinding process, 20 mesh, then 40 mesh and finally 60 mesh. After each grinding step, three FT-NIR spectra were acquired on the vial of sample before the next grinding (smaller sieve opening) step was conducted on it. In this way, for certain variety of biomass, the chemical composition can be considered remaining the same, while the particle size varied. The experimental design is shown in Table 2, while each sample variety followed the procedures shown in Figure 13. Therefore, a major physical variation (particle size) was rationally created, while morphology difference as a minor physical variation was indirectly created at the same time.

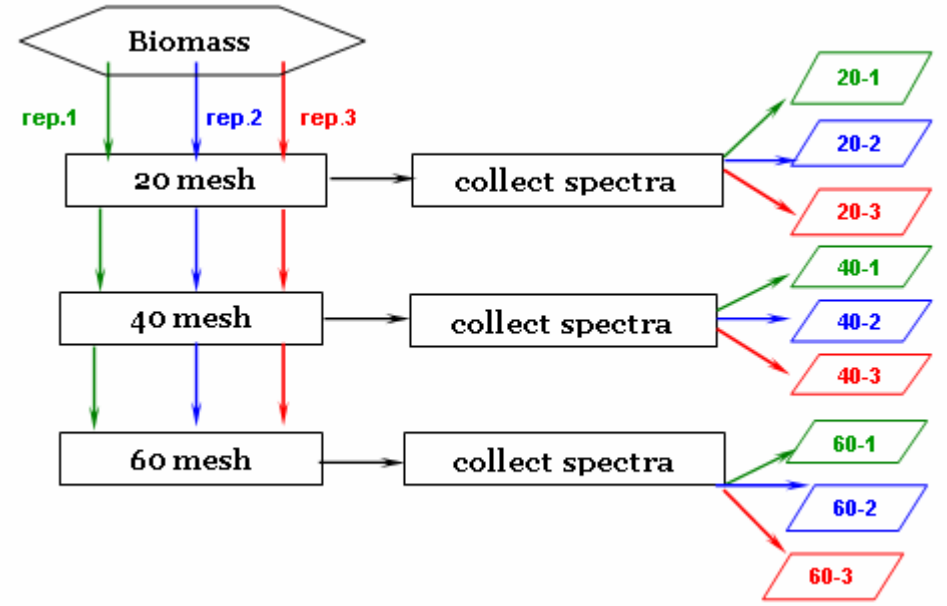


Figure 13 Scheme of experimental procedures for sampling one investigated biomass variety, repeated for *Hu, Hs, G*, respectively

Microscopic imaging was performed on the 40 mesh ground powders of different switchgrass botanic parts. Olympus SZH10 Research Stereo (Microscope) was utilized for the magnification and SONY MAVICA MVC-CD500 was used for taking the image. The amplification of 70× was applied to monitor a single particle and 20× to monitor the particle distribution. Original NIR spectral data were reprocessed using nine different algorithms in the Unscrambler 9.2, and then transported to NCSS 2004 for HCA. The comparison of pretreatment performance was made among the nine different methods, given the knowledge of the chemical and physical characteristics. Nine pretreatment methods of interest were SNV, 1^{st} derivative, 2^{nd} derivative, MSC, EMSC, 1^{st} derivative followed by SNV (1^{st} + SNV), SNV followed by 1^{st} derivative (SNV+1^{st}), 2^{nd} derivative followed by SNV (2^{nd}+SNV), and SNV followed by 2^{nd} derivative (SNV+2^{nd}). Other than the judgment based upon HCA, the study also investigated the pretreatment performance via the predictive results. The original

spectral dataset of switchgrass (Table 3) was pretreated using nine different pretreatment methods, and then modeling was performed between each of the nine spectra dataset and four most important chemical constituents (glucose, xylose, lignin, and ash). Cross-validation was used to evaluate the performance of the nine models based on two statistical criteria (correlation and RMSEP). Larger correlation and lower RMSEP suggested better linearity between spectra data and chemical data.

4.3.2.2.2 Spectral transform algorithm comparison

To determine whether *K-M* or *log(1/R)* provided better biomass NIR modeling, this study assessed it via the experimental point of view. The comparison was performed for switchgrass dataset (Table 3) and corn stover dataset (Table 4), respectively. As spectra were collected in both *K-M* unit and *log(1/R)* unit (Section 4.2.2.2), so for both corn stover and switchgrass, there were two sets of data: *K-M* spectra with chemicals and *log(1/R)* with chemicals. All the other modeling configurations remained identical to ensure a fair comparison. For both corn stover and switchgrass, two sets of models were thus developed, and the selection of the better algorithm was based on correlation and RMSEP calculated via cross-validation.

4.3.2.2.3 Multivariate regression method comparison

Corn stover and switchgrass calibration datasets were utilized for the regression methods comparison. They were considered as two modeling streams and the comparison was preformed thereupon. Three regression methods MLR, PCR and PLS were compared based the corresponded model performance while other modeling configurations (transform algorithm and spectral pretreatment) remained identical,

MLR, PCR, PLS regression were conducted in the Uncrambler 9.2, with spectral data as independent variables and chemical data as the dependent variables. As for MLR, stepwise variable selection was conducted in NCSS to remove the collinearity that highly existed among 1558 spectral variables. The

inclusion significance was set at 0.05 and exclusion significance 0.10, and the selected variables would be used as independent variables for MLR. Correlation and RMSEP from cross-validation were used as the criteria to compare the performance three models and thus determined the best regression method.

4.3.2.3 Model validation

The individual models for corn stover and switchgrass and the general model including both corn stover and switchgrass were thus developed based on the best modeling configurations that determined in Section 4.3.2.2. Leave-one-out cross-validation was used to evaluate the model performance.

Furthermore, an independent validation was performed to further verify the general model prediction performance. This validation dataset included five corn stover and five switchgrass samples (Table 5) that were of different genotype, collected at different time and spatial locations, and not used in the calibration development.

Five wheat straw samples (Table 5) were conducted to investigate the potential of using the developed general model to predict the wheat straw chemical composition. The significance of introducing a third species to the validation is that it will confirm that NIR techniques only probe chemical differences, and the genotype, ecotype and some physical variation can be corrected by the pretreatment method. The general model can be further extended to cover more biomass species as long as the chemical composition is within the calibration region.

Both chemicals and spectral data were collected as the same procedures

Table 5 Information associated with the validation dataset

Varieties	*Reps.*	*Sources*
Corn Stover	5	UT experimental station in Knoxville, TN, 2006 summer
Switchgrass	5	UT experimental station in Milan, TN, 2005
Wheat Straw	5	UT dairy site in Blount county, TN, 2006 summer

as calibration dataset data collection. Afterwards, these fifteen FT-NIR spectra were pretreated by the saved EMSC model (based on the 71 calibration spectra), and then input into the developed general model and calculated the seven constituents' values. The predicted chemical values as well as the measured were transported into a spreadsheet (Excel 2003), and several model evaluation parameters were calculated by applying the corresponding algorithms (Appendix C).

4.3.2.4 HHV analysis

The scanned biomass samples remaining after the usage of wet chemistry were collected by variety and prepared for higher heating value (HHV) testing; therefore, these samples share the same FT-NIR spectra with the samples went through the wet chemistry analysis. The sample preparation was performed in two steps: first, around 0.5 g samples were compressed via a pelletizer, and oven drying to remove moisture. IKA calorimeter system C 200 was utilized to measure HHV, and all HHV was calculated on the dry basis. Measurement was performed in duplicates (Table 6), while three spectral replicates were collected before wet chemistry analysis; this resulted in that each replicate of heating value did not exactly match each individual spectrum.

Table 6 HHV data collection of 15 categories of biomass samples (2 replications)

Category	*Reps.*	*Category*	*Reps.*	*Category*	*Reps.*
Cave-in-rock	2	Leaf-Alamo	2	Sample 1[1]	2
Alamo	2	Leaf-Kanlow	2	Sample 2	2
Kanlow	2	Node-Alamo	2	Sample 3	2
Shelter	2	Node- Kanlow	2	Sample 4	2
NC1-16	2	Internode-Alamo	2		
NC2-16	2	Internode-Kanlow	2		

1. Samples 1-4 are from independent validation dataset.

Therefore, instead of directly using individual HHV, modeling was conducted on the average of two spectral replicates and the average of the three FT-NIR spectra as the independent variables (thus sixteen sets of data) for the PLS regression. All the averaging performance was conducted the Uncrambler 9.2.

CHAPTER V RESULTS AND DISCUSSION

This chapter provides the results corresponded to all the experiments dictated in Chapter IV, along with the discussion and inferences. Section 5.1 presents wet chemistry results (the conventional measurement). Section 5.2 showed how moisture affects the spectra and justified the sampling condition selected in this study. Section 5.3 demonstrates the determination of the best modeling configurations for biomass NIR study, and Section 5.4 presents the validation results for the thus developed models.

5.1 Wet chemistry results

5.1.1 Switchgrass

Among the six switchgrass cultivars, variation was clearly present (Table 7): glucose content ranged from 38.35% in NC1-16 to 46.17% in Cave-in-rock, xylose from 19.65% in cave-in-rock to 22.90% in Shelter, galactose from 2.17% in Cave-in-rock to 3.48% in Alamo, arabinose from 3.40% in NC2-16 to 4.76% in Shelter, mannose from 0.62% in Alamo and Kanlow, lignin from 20.64% in Kanlow to 22.89% in Alamo, and ash ranging from 2.62% in Cave-in-rock to 3.90% in Alamo. The values were calculated by averaging three samplings out of each cultivar. Although Alamo and Kanlow had been reported the highest crop yield, they are observed comparatively low glucose content, while the low-yield upper-land crops Cave-in-rock had larger glucose proportion and indicates higher ethanol production efficiency. Therefore, the overall economical efficiency, considering both crop productivity and ethanol conversion efficiency should be studied to finally draw a conclusion to guide planting. Furthermore, an interesting finding lied in NC2-16, it had high glucose content, low ash and lignin content, which met the criteria for good feedstock for bio-ethanol production; moreover, its lowland ecotype characteristics suggest that it has high yield. Therefore, the results showed that this cultivar is promising and worthy of more attention. The variability was enhanced with the additional samples from the manually-separated botanic fractions: for glucose, the lower boundary was brought down

Table 7 The chemical composition of different switchgrass cultivars and botanic fractions

	Glucose	Xylose	Galactose	Arabinose	Mannose	Lignin	Ash
Cave-in-rock	46.17[1]±.36[2]	19.65 ±.15	2.17±.05	3.65 ±.56	0.94 ±.13	21.73 ±.57	2.62 ±.04
Alamo	40.41 ±.28	22.90 ±.73	3.48 ±.27	4.19 ±.36	0.62 ±.08	22.89 ±..37	3.90 ±.19
Kanlow	40.34 ±.31	20.71 ±.06	3.34 ±.16	4.11 ±.32	0.62 ±.05	20.64 ±..01	3.21 ±.07
Shelter	41.53 ±.39	23.76 ±.20	3.11 ±.10	4.76 ±.11	0.97 ±.08	21.91 ±.25	2.64 ±.25
NC1-16	38.35 ±.68	22.64 ±.95	2.82 ±.24	3.62 ±.13	1.28 ±.31	22.83 ±.55	3.29 ±.23
NC2-16	44.71 ±.51	21.64 ±.70	2.61 ±.12	3.40 ±.05	1.11 ±.22	21.49 ±.57	3.26 ±.16
Leaf-Alamo	38.59 ±.22	19.58 ±.22	3.56 ±.13	4.75 ±.15	1.10 ±.14	24.43 ±.33	4.90 ±.08
Leaf-Kanlow	40.45 ±.17	23.98 ±.17	1.87 ±.42	3.35 ±.53	1.07 ±.02	25.10 ±.13	3.74 ±.01
Node-Alamo	39.09 ±.27	25.87 ±.27	2.56 ±.13	5.39 ±.11	1.00 ±.11	21.43 ±.12	2.62 ±.12
Node-Kanlow	37.81 ±.20	25.69 ±.20	2.15 ±.12	4.59 ±.02	0.67 ±.11	22.18 ±.24	1.49 ±.28
Internode-Alamo	45.54 ±.50	24.35 ±.50	2.15 ±.29	3.17 ±.22	0.70 ±.05	21.43 ±.21	2.19 ±.44
Internode-Kanlow	43.45 ±.22	25.76 ±.22	2.01 ±.12	3.60 ±.03	1.15 ±.12	21.36 ±.12	1.60 ±.07

1. Mean was calculated based on the three replicates of each category (in %w/w).
2. Standard error was calculated based on the three replicates of each variety (in %w/w).

to 37.81% by node (Kanlow); both ends were extended for xylose, downwards by leaf (Alamo) to 19.58% and upwards to 25.87% by node(Alamo); the range of galactose was enlarged both ends to 1.87%-3.56% and ash's range as well (1.60%-4.90%); and for lignin, the higher range was achieved to 25.10% by leaf (Kanlow). As most of the extremes are achieved by the fractions, the issues of having lower valuable constituent can be compensated for the high-yield plants-Alamo and Kanlow. The chemical differences were observed via different botanic fractions, which indicated the efficiency and significance of using certain botanic fractions instead of the whole plot to enhance the bio-conversion. The trend can be observed that internodes have higher glucose content than leaves and nodes, while leaves have higher lignin and ash contents yet lower xylose content. This suggests that it is very positive that bio-ethanol production increases its efficiency by 10% via utilizing prescreened internodes than the whole switchgrass plot, which is mechanically feasible (Klasek, 2006). Also, the observation that botanic parts provide larger variability than several cultivars indicates that using botanic

parts is an effective alternative to create variability for model calibration.

5.1.2 Corn stover

Large variability was observed (Table 8) among different corn stover botanic parts, which again proved that this manual separation can efficiently provide large variability for multivariate analysis. Glucose content was ranging from 32.39% in Node to 44.03% in sheath, xylose from 18.10% in nodes to 25.26% in husks, galactose from 1.69% in rinds to 2.73% in husks, arabinose from 2.63% in rinds to 5.42% in husks, mannose from 0.93% in pith to 1.77% in husks, lignin from 16.14% in husk to 23.95% in leaves and ash ranging from 2.42% in husks to 8.79% in leaves. Generally, the internodal part, composed of the sheath, rind and pith, was preferable for fermentation than node and leaves, since it contained significantly higher glucose content, comparably lower lignin content. Moreover, husk was the most valuable for the bio-ethanol production, since 1) its overall sugars contents were comparatively high while previous researches (Van Zyl. et al., 1988; Alterthum and Ingram, 1989; Bothast et al., 1994; etc) proved that both cellulose and hemicellulose can be converted to ethanol, and 2) also husk had the lowest lignin and ash content. Since compared to separating corn leaves, nodes, internodes, it is much easier to separate husk from cob (Kracl, 1986) which is also necessary for food supply. So it is promising

Table 8 The chemical composition of different botanic parts of corn stover

	Glucose	Xylose	Galactose	Arabinose	Mannose	Lignin	Ash
Node	32.39[1] ±1.122[2]	18.10 ±1.05	2.41 ±.25	4.39 ±.21	1.12 ±.17	23.6 ±.41	3.7 ±.28
Pith	43.37 ±1.49	19.66 ±.88	1.85 ±.19	3.29 ±.38	0.93 ±.31	19.88 ±.63	4.86 ±1.00
Sheath	44.03 ±.85	19.44 ±.94	2.27 ±.48	4.5 ±.64	1.43 ±.48	16.31 ±6.02	5.42 ±1.15
Rind	42.04 ±.63	18.91 ±1.32	1.69 ±.35	2.63 ±.58	1.18 ±.31	22.68 ±1.71	3.8 ±.46
Leaf	34.20 ±2.08	18.31 ±1.28	2.49 ±.57	3.52 ±.27	1.32 ±.10	23.95 ±.46	8.79 ±4.24
Husk	41.78 ±2.87	25.26 ±4.87	2.73 ±.31	5.42 ±.72	1.77 ±1.03	16.14 ±1.79	2.42 ±1.19
Whole[3]	36.86 ±1.90	21.52 ±1.96	2.41 ±.52	3.55 ±1.09	1.25 ±.41	22.1 ±2.25	2.37 ±.38

1. Mean is calculated based on the three replicates of each category (in %w/w).
2. Standard error is calculated based on the three replicates of each category (in %w/w).
3. Whole corn stover, including node, pith, sheath, rind and leaf.

and meaningful that husk is favorable for bio-ethanol production. The result also indicated that leaves had high lignin proportion thus good for bio-power plants but the fouling problems need to be paid attention to, due to its high ash content.

5.1.3 Chemical composition overview

The variability that exhibits among the botanic parts of just one corn stover cultivar is much greater than the variability created by different switchgrass cultivars and even their botanic parts. This observation implies that the variation lies in the corn stover species, generally, is larger than switchgrass species. Consequently, when switchgrass is used as the feedstock, less economic concern would arise during long-term operation, since its quality is comparably stable and consistent; on the contrary, when corn stover as feedstock, due to its large variability, close attention should be paid to monitoring the feedstock composition.

Furthermore, corn stover and switchgrass show the consistency in that internodal parts have higher glucose content than nodes and leaves, and leaves tend to have more lignin content, suggesting that these tendencies might exist generally among biomass species.

5.1.4 Higher heating value for switchgrass dataset

Figure 14 presents the HHV results for all the switchgrass varieties, following the procedure dictated in Section 4.3.3.4. Among the six switchgrass cultivars, NC1-16 had very high heating value, thus preferable for co-firing and combustion; and Alamo and Kanlow were also promising for bio-power industry considering their high yields and their moderately large HHV. As for the botanic parts, leaves had significantly larger HHV than node and internodes. Compared with the chemical properties of switchgrass presented earlier in Table 7, it is found that the samples with higher lignin content had comparatively higher HHV, while those with higher glucose proportion exhibited comparatively lower HHV. It was reasonable because cellulose had the lowest heating value (11.7 kJ/g) and lignin had comparatively high heating value (24.1 kJ/g) (Raveendran and Ganesh,

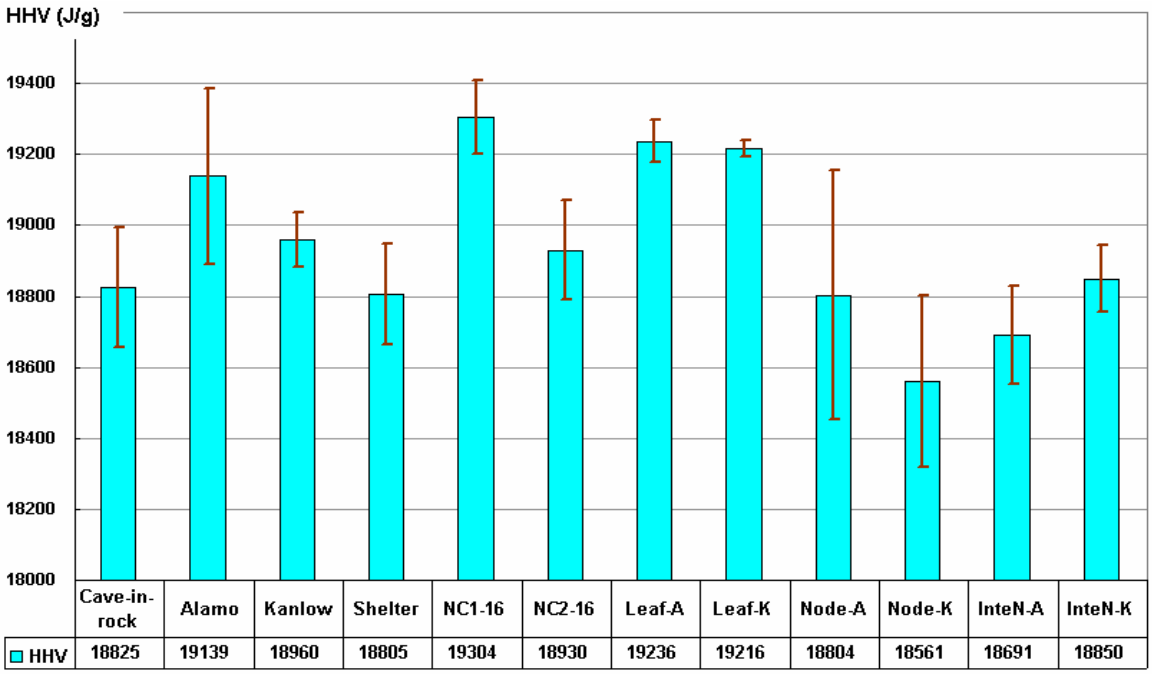


Figure 14 HHV of different switchgrass cultivars and botanic parts [1,2,3]

1. The ordinate scale starts at 18200 J/g.
2. The listed HHV values are calculated from two measurements for each category.
3. The error bar shows ±standard error, calculated from the two measurements.

1996). Also, large HHV variation exhibited within each investigated variety with the standard error ranging from 22.67 to as large as 351.43 J/g.

5.2 NIR Water bands and justification of using dry samples

All the observations were derived from the same sample in a manner of increasing moisture content (Section 4.3.1); therefore, they could be assumed to have the same proportion of all the chemical constituents with only moisture content variation. In the spectra of the series of samplings, it is very clear moisture content significantly affected to the entire NIR spectra (Figure 15), since the response of O-H stretching to NIR is overwhelming. The significant increase of absorption occurred in the two regions (as marked on Figure15), especially region a .

Principle component analysis was performed to confirm this direct visual

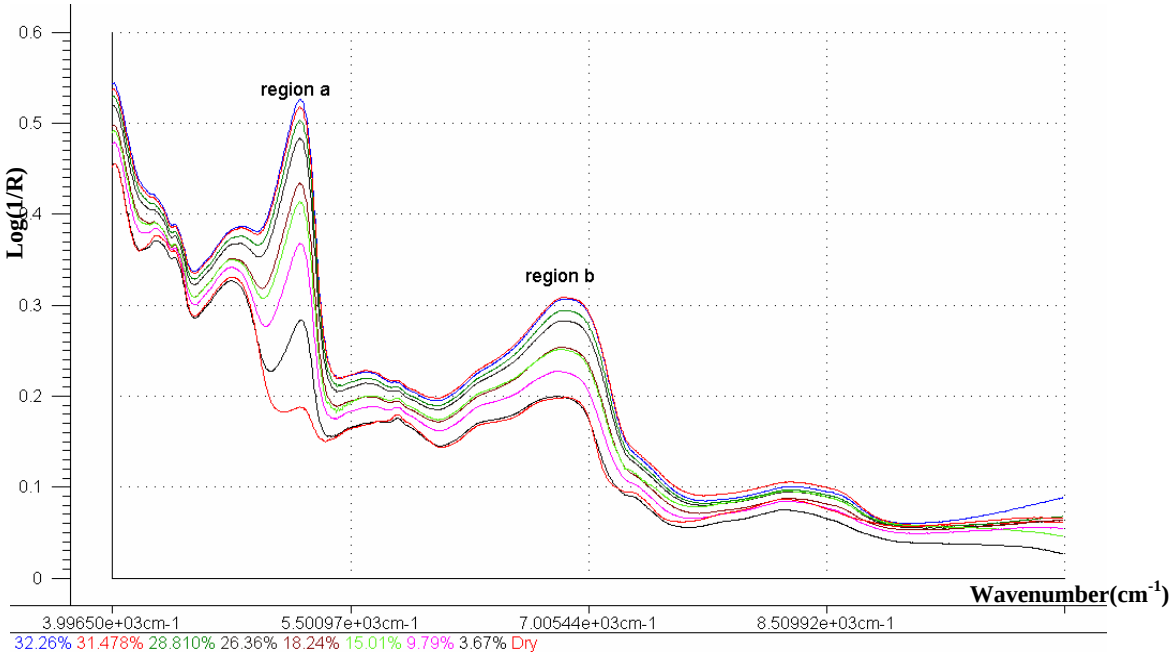


Figure 15 The FT-NIR spectra of all the moisture samplings. From the bottom curve of dry sample (0% moisture content) up, the spectra correspond to moisture content of 3.67%, 9.79%, 15.01%, 18.24%, 26.36%, 28.81%, 31.48%, 32.26%, respectively. Moisture change caused significant spectral variation in region a, and region b as well.

judgment. It was very clear that observations with the increased moisture content were distributed following PC1 direction, which accounted for 97% of the total variance of this dataset (Figure 16). Therefore, PC1 was determined to reflect moisture content and its associated loading plot (Figure 17) was confirmed to carry the characteristic bands information attributed to moisture, or water.

The result in Figure 17 showed the agreement with some former studies (Williams, 1992; Neimanis et al., 1999, Rantanen et al., 2000),, an intense peak appears around 5180 cm^{-1}, and consequently this region was assigned to the bond of –OH stretching due to water, covering from 4800 cm^{-1} and 5450 cm^{-1}. Besides, the broad peak around 7067 cm^{-1} as well exhibited less yet still significant correlation with water. Thus the region between 6804 cm^{-1} and 7167 cm^{-1} was significantly affected by water as well. The spectral curve presentation between 4800 cm^{-1} and 5450 cm^{-1} can be used as a flag to on-time signal whether the sample is dry or wet.

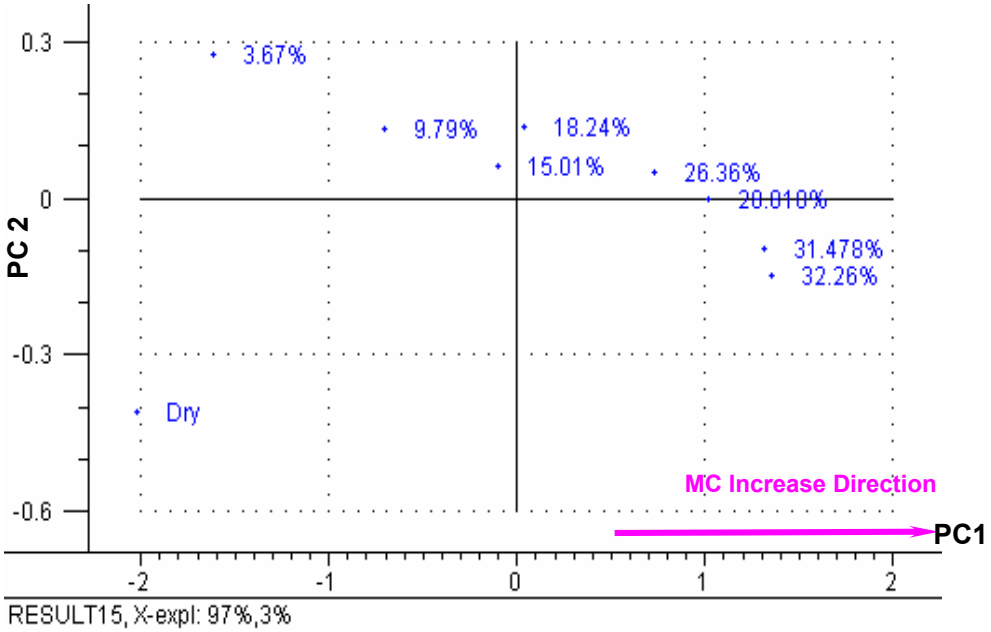


Figure 16 The score plot of PCA applying to the 9 spectra of different moisture samplings depicted in Figure 15

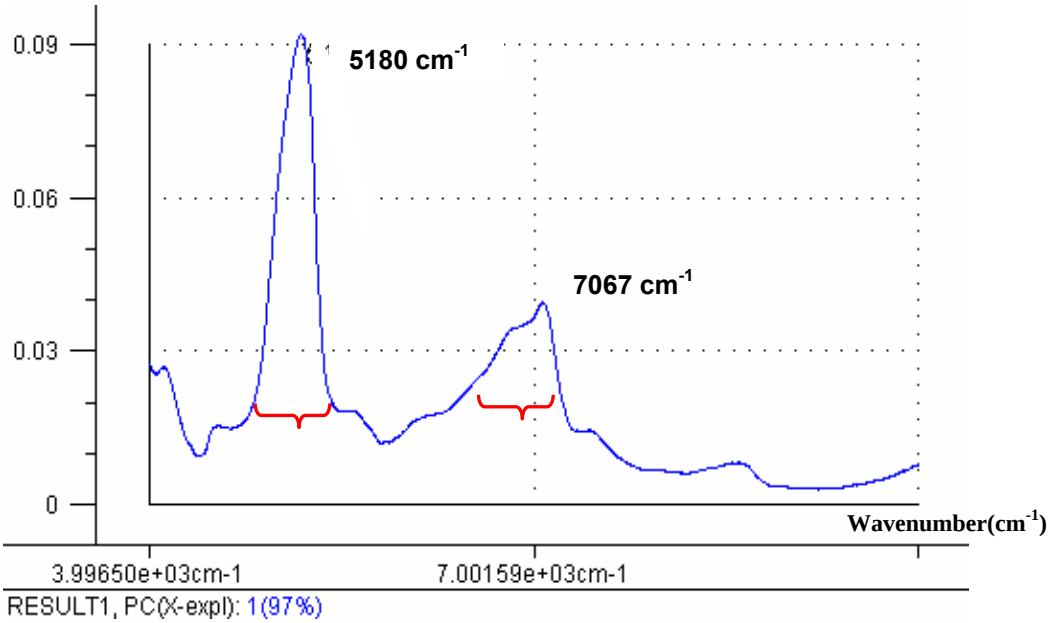


Figure 17 The loading plot associated with PC1 depicted in Figure 16

This study also found that the integrated peak area of the spectra in the assigned regions was linearly correlated with moisture content (Figure 18). The integration was performed in the Varian Resolution Pro software with the forced cut-off edges (4800 cm^{-1} and 5450 cm^{-1}, 6804 cm^{-1} and 7167 cm^{-1}, respectively), and linearly regressed against moisture content. The data points were closely distributed around the linear regression line with R^2 over 0.99 for both regions, which demonstrated that moisture content and spectral integrated area were highly correlated in these two regions. Therefore, the wavenumbers 4800 cm^{-1}~ 5450 cm^{-1} and 6804 cm^{-1}~7167 cm^{-1} were proved as water or moisture bands.

Water, essentially the O-H bonds, has a significant influence to the entire NIR spectra, and this overwhelming effect may interfere with the modeling work

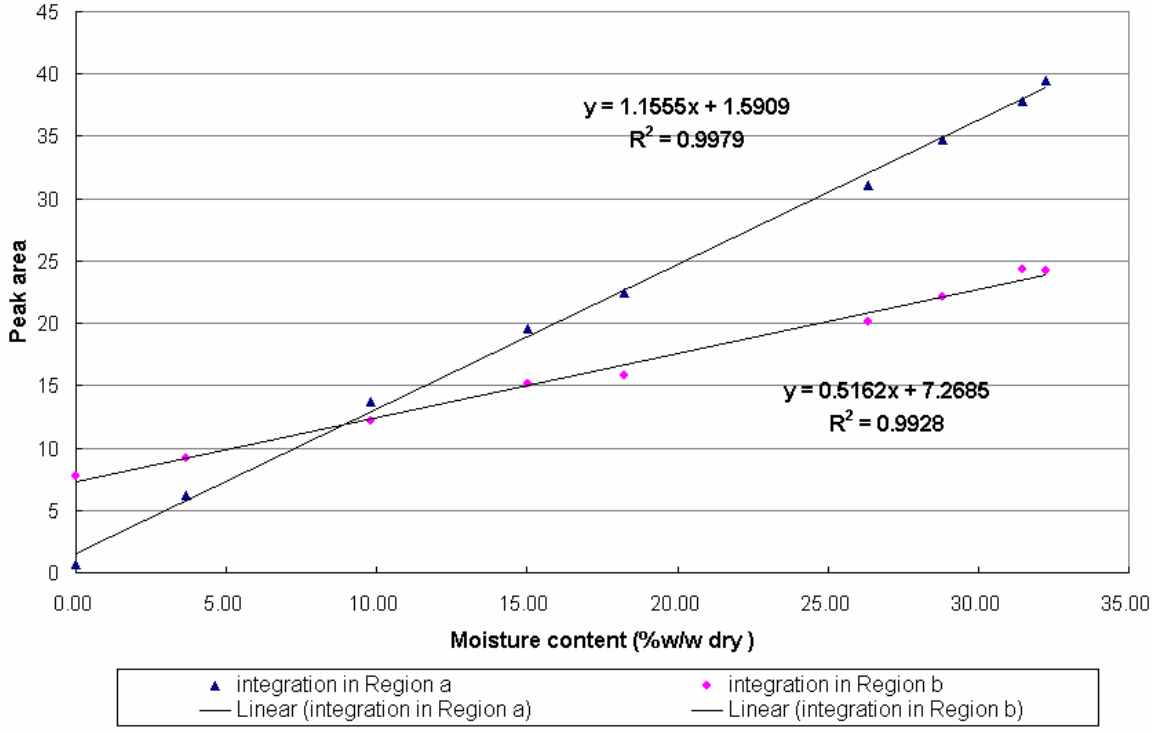


Figure 18 The plot of the integrated peak (two) area vs. moisture content. The line with triangles presents the relationship between the integrated area and moisture content in the most significant region a (4800 cm^{-1} and 5450 cm^{-1}); the line with dots presents the relationship in the second significant region b (6804 cm^{-1} and 7167 cm^{-1}).

on the other chemical properties by the even slight variation of sample moisture content. Furthermore, in the two assigned regions, other than O-H in the moisture, O-H stretching also exists in the other compositional constituents such as carbohydrates, and other bonds may have absorbance in these two regions (especially 6804 cm^{-1} and 7167 cm^{-1}) as well. In other words, the existence of moisture would overwhelm the variations in these two regions attributed to other compounds. Furthermore, since wet chemistry analysis requires oven-dried samples at the start point, acquiring FT-NIR samples on the dried sample ensured that the spectra exactly matched wet chemistry measurement since oven-drying might cause some chemical loss. These reasons justified why the FT-NIR spectra were collected on dried samples, since the major task of fast NIR analysis was to probe the compositional chemicals, not moisture content.

5.3 The development of best modeling configurations

5.3.1 Spectral pretreatment

5.3.1.1 The necessity of spectral pretreatment

The physical characteristic is a critical source of interference in investigating of chemical properties. However, how to differentiate the chemical and physical causes and claim that the data pretreatment is removing the essentially physical interference is not simple. This study addresses this pretreatment issue from several perspectives:

5.3.1.1.1 Microscopic imaging results

The particle distribution of the leaves of Alamo was clearly presented through microscopic image (Section 4.3.2.2.1). Although all the particles passed through the same sieve with the opening diameter of 0.425mm, the particle size variation was clearly observed (Figure 19). The morphological difference lied in different botanic parts correlated to botanical structure which could be observed under microscope (Figure 20). Node particles usually had chunky shapes; while internode and leaf particles mostly existed as long tissues (leaf particles also had flat profile). Their surface texture exhibits differences as well.

Figure 19 Microscope view of particles of Alamo-leaf sample after 40 mesh grinding

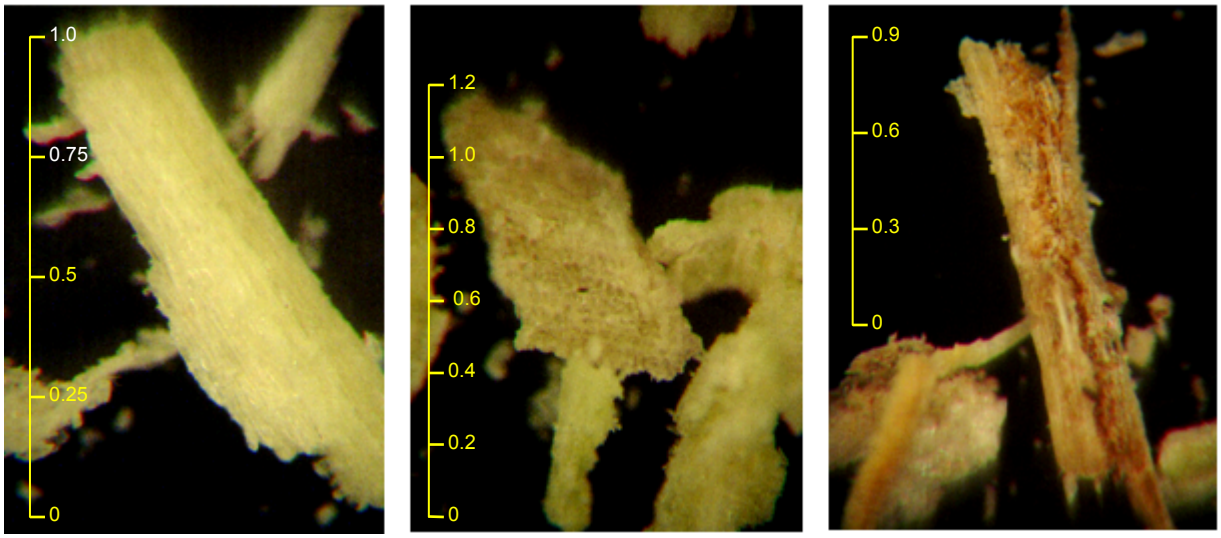


Figure 20 Micrographs of the ground (through 40 mesh) particles of Alamo botanic parts. Internodes (left), Nodes (middle), Leaves (right). (All the values in the scale is in mm)

All these graphs clearly demonstrated that to investigate biomass, it was very difficult and almost impossible to achieve ideally homogenized NIR sampling (all the particles had identical physical properties, such as size, shape, and texture).

5.3.1.1.2 Original spectral analysis

The original spectra obtained via the particle size control experiment (Section 4.3.2.2.1) scattered greatly (Figure 21), which suggested that the physical effect due to particle size contributed significantly to the overall variation. One argument is that many spectra had larger signal intensity than some others throughout all the wavenumbers, which was impossible because the biomass samples had the same constituents. Statistical analysis was employed to demonstrate that the particle size effects disguised the variation attributed to chemical composition difference.

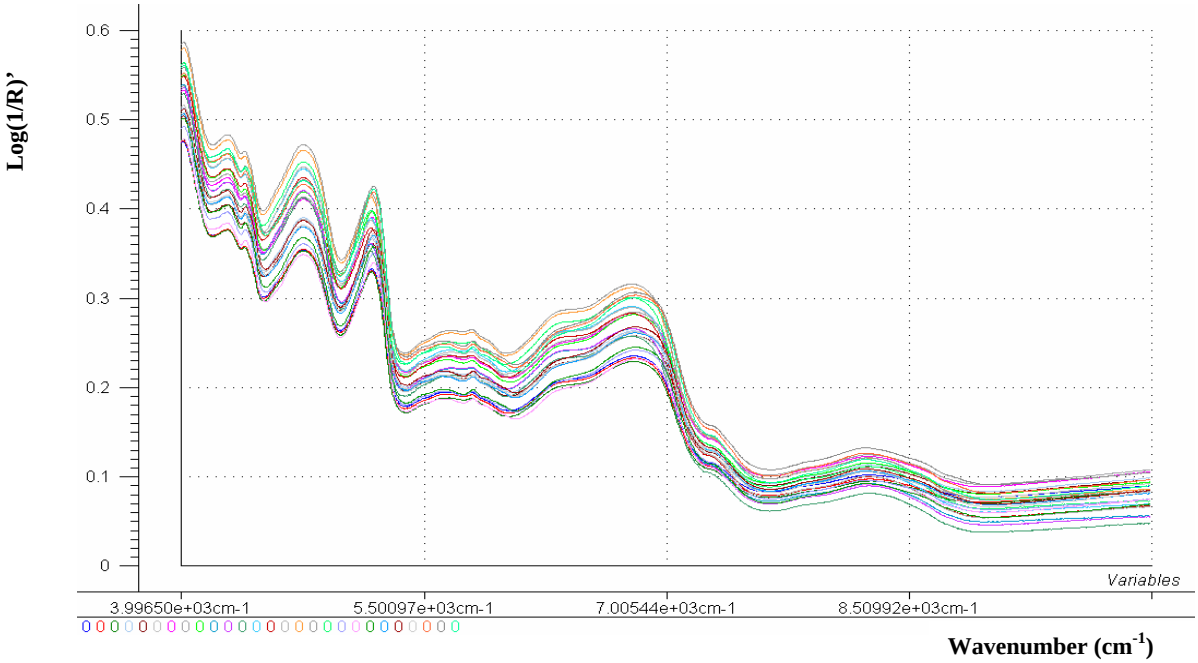


Figure 21 The 27 original spectra, 3 repetitions for each of 3 particle size for 3 biomass varieties.

First, PCA was conducted on these spectra dataset, all wavenumber variables were utilized. Along PC1 direction in the score plot (Figure 22), the tendency of particle size distribution was evident: the spectral observations with finer particles were distributed towards the left side, and larger particles toward right and this PC explained 82% of the total variance of the spectra dataset. Meanwhile, from the score plot, along PC2 direction, the sample distribution tends to reflect certain chemical characteristics: Hr and G observations exhibited on the top and Hs ones appeared at lower level. However, even if completely contributed by chemical difference, the explained variance was much lower than the explained portion by particle size difference.

Compared particle size distribution along PC1 in Figure 22 with the PC1 associated loading plot (Figure 23), the larger particle samples had larger NIR absorption across the entire NIR region, since the loading values remained positive for all the wavenumbers. Also the loading curve shape had great resemblance to the original spectra, which suggested given the same amount of NIR incident beam, diffuse reflectance happened more times in the sample with larger particles. It is reasonable because with large space between particles, NIR beam was easier to get through the surface and hit more inner particles, thus more absorption took places. Furthermore, this suggested that other than an offset term, particle size variation also brought in the multiplicative term (refer to Section 2.3.1). The dendrogram (Figure 24) states the same problem from another perspective. The major three clusters were expected to represent the three biomass varieties, since the target of using NIR was to probe chemical difference, that is to say, the three major clusters (marked as A, B, C) should represent G samples, Hs samples, and Hr samples, since G, Hs, and Hr represented three varieties (Section 4.3.2.2.1). However, the cluster A was composed solely of the samples after 60 mesh and cluster B and C reflected more particle size variation than variety information. This hierarchy structure was far from the expectation that three major clusters composed of G, Hs and Hr, respectively.

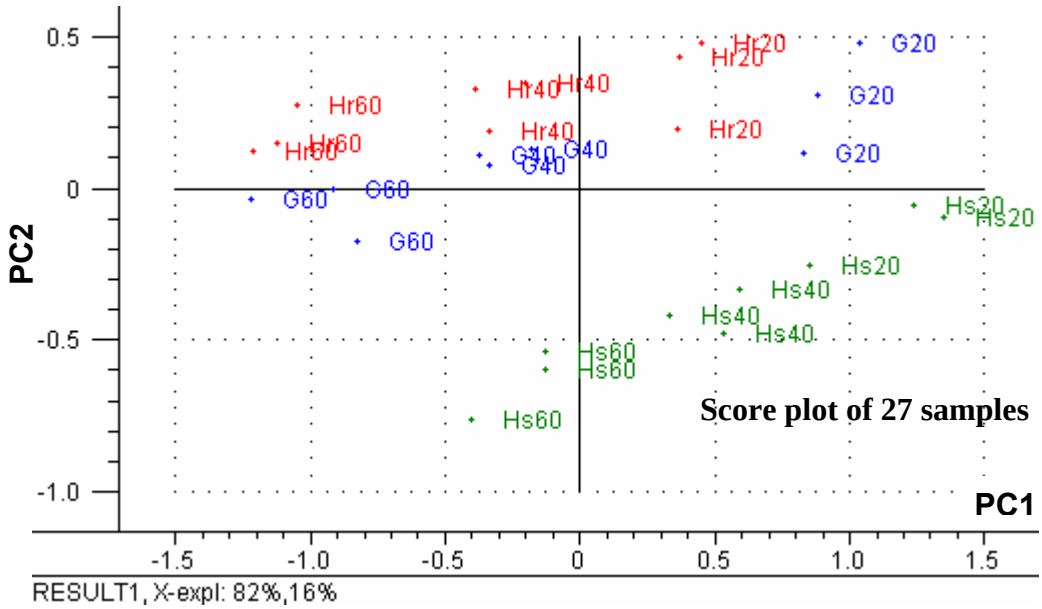


Figure 22 PCA score of 27 samples: PC1 vs. PC2. Three biomass varieties: Hr: Husk from the regular corn cultivar DeKalb DK64-10RR; Hs: Husk from the sweet corn cultivar Incredable; G: straw of switchgrass cultivar Shelter. 20, 40, 60 stand for the grind mesh size, thus biomass particle sizes decrease as the number increases.

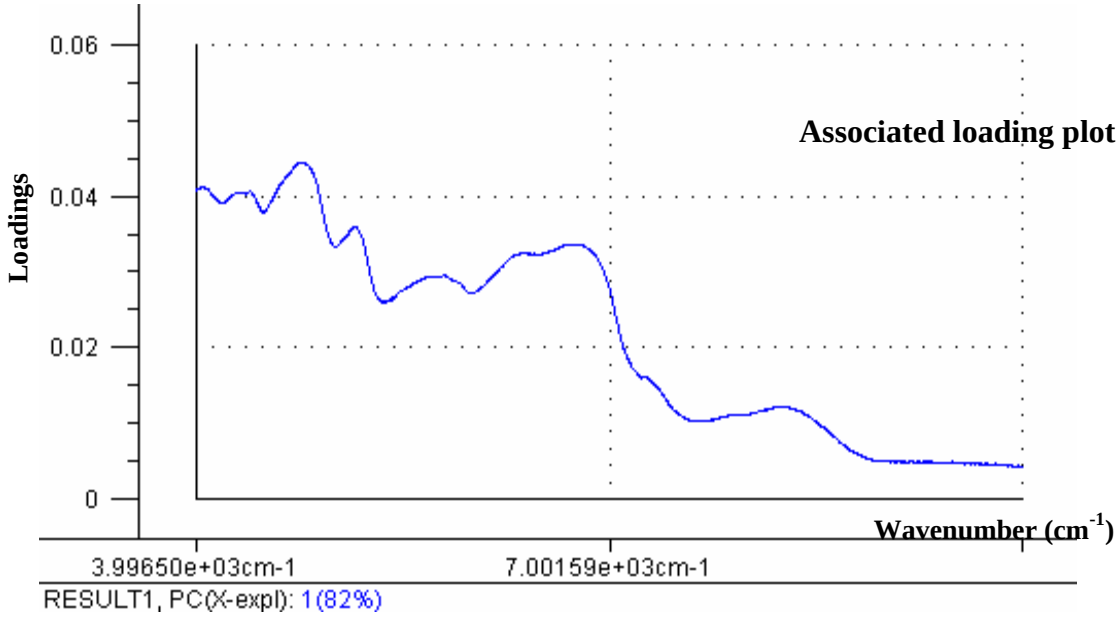


Figure 23 PCA loading plot associated with PC1 displayed in Figure 22

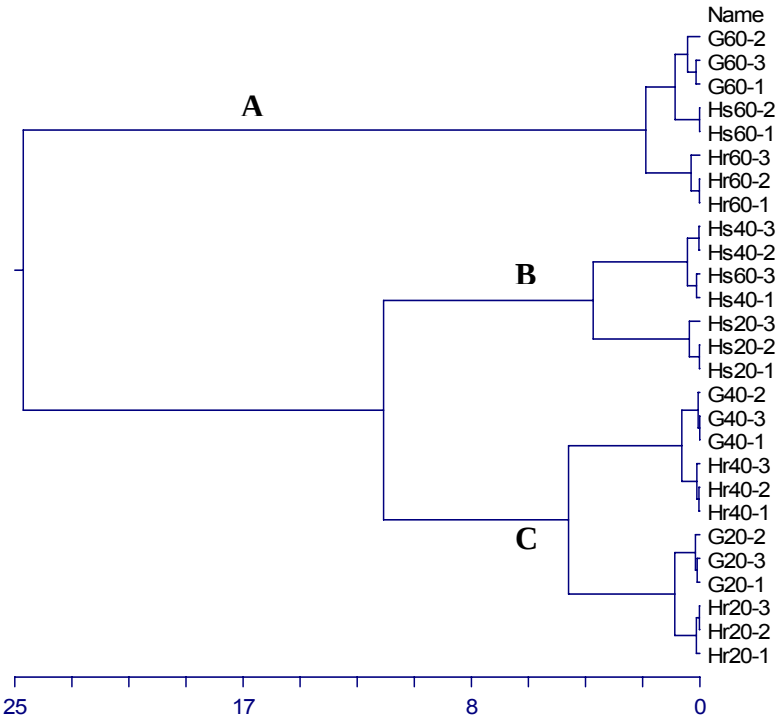


Figure 24 The dendrogram of HCA of the 27 original FT-NIR spectra

5.3.1.2 Pretreatment selection

5.3.1.2.1 Hierarchical clustering results

HCA was applied in the same manner on the datasets pretreated by all the nine methods described in Section 4.3.2.2.1. For all the dendrograms (Figure 25), the three major clusters clearly represented sample varieties, which indicated that physical interferences had been reduced and chemical characteristics were thereby enhanced. These structures again demonstrated the significance of spectral pretreatment in FT-NIR analysis of biomass chemical composition.

The efficiency in reducing physical variation varied amongst the nine methods; and this point can be reached by examining the dissimilarity of clusters. The dissimilarity of clusters can be read from the horizontal position of the split. Compared to the other eight methods, EMSC exhibited the best performance for

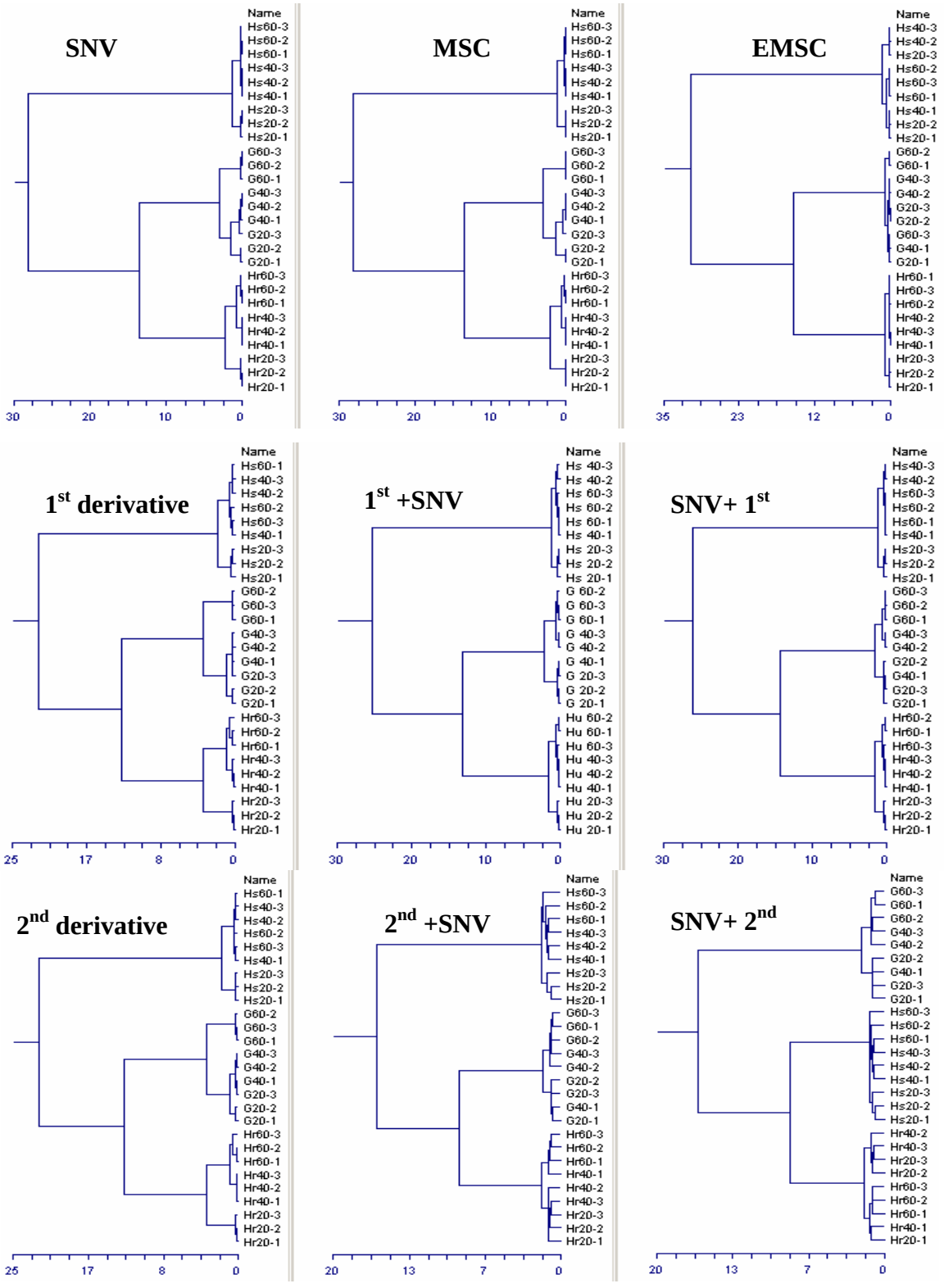


Figure 25 The dendrograms of the 27 FT-NIR spectra after different pretreatments

Table 9 The dissimilarity distances calculated during the last three mergence in HCA for different pretreatments. 1st row: the dissimilarity distances of the last merged two clusters; 2nd row: the dissimilarity distances of the last 2nd merged two clusters; 3rd row: the dissimilarity distances of the last 3rd merged two clusters, which also represent the largest dissimilarity attributed to particle size

	SNV	*MSC*	*EMSC*	*1st derivative*	*1st+SNV*	*SNV+1st*	*2nd derivative*	*2nd+SNV*	*SNV+2nd*
1	28.2	28.2	30.7	22.1	25.4	26.2	22.1	16.1	16.4
2	13.4	13.5	15.0	12.8	13.2	14.3	12.8	8.9	8.3
3	3.02	3.04	1.34	3.60	2.18	1.61	3.60	1.75	2.04

reducing physical interference, based on the observation that the ratio of the dissimilarity distance resulting from chemical compositions to that from physical effects is maximized. From the scale of the dissimilarity axis, it is clear that EMSC provided larger distance between the three major clusters.

The quantified comparison is further presented in Table 9. The first two rows represent the difference among the three major clusters, or sample varieties, which is the larger the better; and the 3rd row is the largest dissimilarity attributed to particle size differences, the smaller the better. EMSC maximized the dissimilarity attributed to chemical variations, and minimized the dissimilarity attributed to physical variation. EMSC quantitatively generated the best pretreatment result for subsequent investigation of chemical composition.

Compared to the original spectra (Figure 21), it was evident that large variation due to physical interferences had been removed (Figure 26), which again proved that physical variation was significant and should be removed; however, only with the reasoning above, it was safe to announce that the removed variation was attributed to physical effects.

5.3.1.2.2 Prediction result comparison

Last section used HCA to demonstrate EMSC as the best pretreatment method from qualitative angle, as the spectral grouping presentation after EMSC pretreatment best reflected the chemical composition differences. The discussion

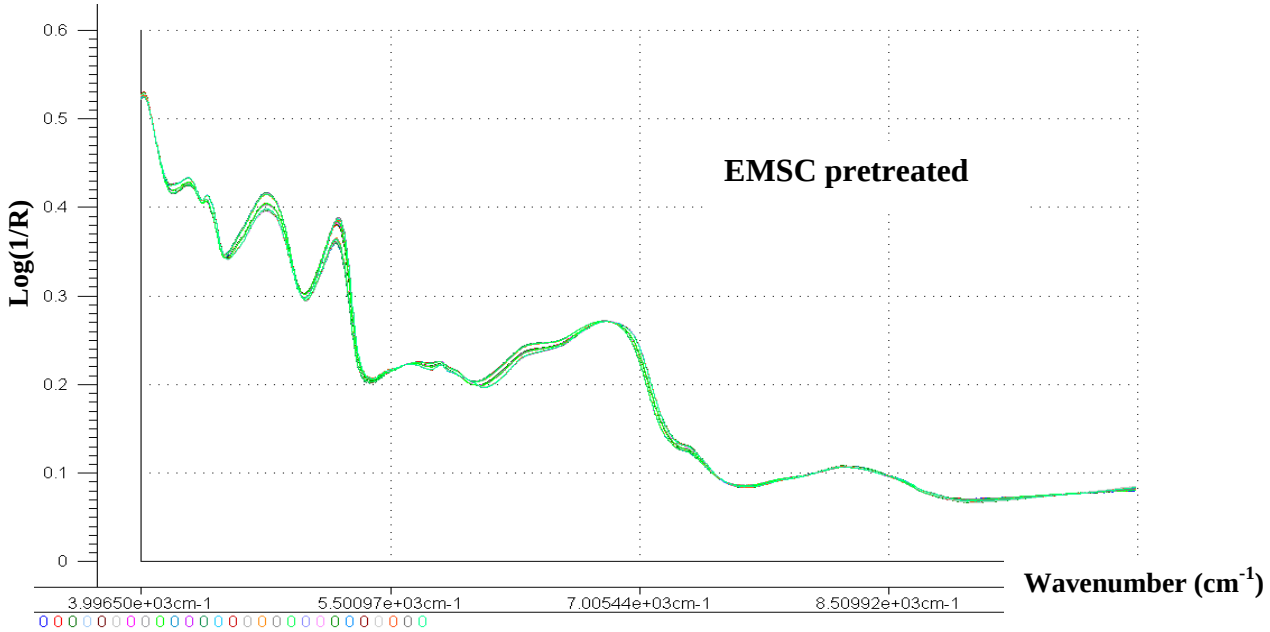


Figure 26 Spectral presentation after applying EMSC pretreatment on the original spectra (Figure 21) of switchgrass samples

in this section focuses on the quantitative perspective based on the cross-validation result of the nine models derived from differently pretreated spectral data (Section 4.3.2.2.1). Two parameters were employed as the criteria to define a good model: 1) Correlation (Table 10) was used for assessing the model fitness and it represents better model when it is closer to 1. 2) RMSEP (Table 11) was used to evaluate model prediction performances and the model performs more accurately if it has smaller RMSEP. Since glucose takes up the largest portion in the biomass chemical composition (averaged 41.37 %w/w) and is a significant source for the fermentation and ethanol production, the glucose content is the most important factor that judgment should be based upon. Clearly, EMSC pretreatment provided best model for glucose, since the associated model had highest correlation and lowest RMSEP. Other than glucose, EMSC associated model had the lowest RMSEP for predicting ash,

Table 10 Correlation[1] comparison of switchgrass models applied with different pretreatments [2]

Pretreatment	*Glucose*	*Xylose*	*Lignin*	*Ash*
Original	0.901	0.882	0.743	0.946
SNV	0.955	0.935	0.861	0.955[4]
1st derivative	0.942	0.942	0.871	0.949
2nd derivative	0.934	0.887	0.849	0.928
EMSC	0.970[3]	0.944[3]	0.902[4]	0.956[3]
1st +SNV	0.951	0.931	0.909[3]	0.945
SNV+1st	0.956[4]	0.937	0.897	0.951
2nd +SNV	0.948	0.935	0.846	0.905
SNV+2nd	0.942	0.942	0.824	0.917
MSC	0.943	0.943[4]	0.856	0.950

1. Larger correlation suggests better model.
2. Original: The model was developed using original spectra without any pretreatment; its correlation is much poorer than all the pretreated ones. SNV: standard normal variate. EMSC: extended multiplicative signal correction. MSC: multiplicative scatter correction. 1st: 1st derivative. 2nd: 2nd derivative.
3. The largest correlation in each column
4. The second largest correlation in each column

Table 11 RMSEP[1] comparison of switchgrass models applied with different pretreatments [2]

Pretreatment	*Glucose (%w/w)*	*Xylose (%w/w)*	*Lignin (%w/w)*	*Ash (%w/w)*
Original	1.250	1.042	1.178	0.309
SNV	0.819	0.782	0.670	0.282[4]
1st derivative	0.954	0.739	0.650	0.299
2nd derivative	1.000	1.018	0.693	0.353
EMSC	0.683[3]	0.730[4]	0.570[4]	0.278[3]
1st +SNV	0.798	0.805	0.517[3]	0.310
SNV+1st	0.780[4]	0.640[3]	0.579	0.285
2nd +SNV	0.906	0.785	0.698	0.405
SNV+2nd	0.94	0.738	0.746	0.378
MSC	0.809	0.734	0.598	0.290

1. Root mean error for prediction: smaller RMSEP suggests more accuracy of model prediction, since it means the predicted chemical content is closer to the measured value.
2. See Table 10 for the abbreviations.
3. The smallest RMSEP in each column.
4. The second smallest RMSEP in each column.

ash, and it had the second lowest RMSEP for lignin and xylose, which are the other two major constituents next to glucose. EMSC also provided largest correlation for xylose and ash, and the 2nd largest correlation for lignin with a slight difference to the largest one provided by 1st +SNV. Although the combination of 1st derivative and SNV exhibited less than EMSC yet still good performance, one concern may be raised when the pretreatment involves derivative algorithm. Derivatives usually enlarge the noise, therefore, if the developed model is transferred to other spectrometer (lower signal-to-noise ratio than FT-NIR spectrometer), problems will occur. Therefore, overall, EMSC turned out to be the best pretreatment method for biomass composition analysis, based on model prediction results.

The result also demonstrated the quantified evident that the model developed on the original spectra did not perform well especially for the three major constituents while pretreatments improved the model accuracy.

5.3.2 Spectral transform algorithm selection

As described in Section 4.3.2.2.2, there were two sets of models developed for corn stover and switchgrass, respectively. Correlation and RMSEP corresponding to the two model sets (the one derived from *K-M* spectra and the one derived from *log(1/R)* are presented in Table 12. For both corn stover and switchgrass dataset, the model derived from the *log(1/R)* spectra had larger correlation and smaller RMSEP than their counterparts derived from the *K-M* spectra for all the seven chemical analytes. This consistency for both biomass species demonstrated that *log(1/R)* transform algorithm provided better linearity between chemical information of biomass samples and their FT-NIR spectra. In addition, in case of any change of the reference reflector (gold plate in this study) in the later-on stage, such as replacement, *log(1/R)* transform provides a simpler systematic adjustment than *K-M* transform.

Because the change of the reference reflector may result in the change of background spectra (I_o at each wavenumber changes), and consequently affect

Table 12 Cross-validation results comparison between *K-M* and *log(1/R)*

		Glucose	*Xylose*	*Galactose*	*Arabinose*	*Mannose*	*Lignin*	*Ash*
				Corn stover [1]				
Correlation	*K-M* [2]	0.911	0.838	0.640	0.732	0.709	0.800	0.882
	Log(1/R)[3]	0.947	0.901	0.838	0.953	0.765	0.962	0.918
RMSEP (%w/w)	*K-M*	1.887	1.687	0.379	0.694	0.385	2.366	1.214
	Log(1/R)	1.407	1.346	0.201	0.341	0.321	1.087	0.700
				Switchgrass [4]				
Correlation	*K-M*	0.949	0.918	0.835	0.819	0.522	0.884	0.945
	Log(1/R)	0.975	0.960	0.925	0.855	0.674	0.939	0.953
RMSEP (%w/w)	*K-M*	0.900	0.874	0.333	0.413	0.224	0.622	0.310
	Log(1/R)	0.633	0.620	0.235	0.374	0.203	0.458	0.266

1. Modeling results on corn stover calibration dataset.
2. The model is developed using the *K-M* transformed NIR spectra. PLS regression is applied; correlations and RMSEPs are generated from cross-validation.
3. The model is developed using the *log(1/R)* transformed NIR spectra. PLS regression is applied; correlations and RMSEPs are generated from cross-validation.
4. Modeling results on switchgrass calibration dataset.

the reflectance at each wavenumber because of the equation $R=I_a/I_o$. So with *log(1/R)* transformed spectra, the systematic bias is a constant across the entire NIR region and can be adjusted easily to the new situation from the formerly developed model. While this simplicity does not occur on the *K-M* transformed ones, a completely new calibration model needs to be developed. Therefore, *log(1/R)* transform algorithm was determined in FT-NIR analyses of biomass chemical composition.

5.3.3 Regression method determination

The results for the comparison of three multivariate regression methods (Section 4.3.2.2.3) are discussed. MLR method was first studied. Stepwise regression on the switchgrass dataset generated seven series of variables (Table 13) that were significant for the predictions of the seven chemical analytes. Likewise, another seven series of significant variables were shown in

Table 13 Stepwise variable selection results on corn stover and switchgrass calibration dataset respectively [1,2]

Chemicals	Significant variables (cm^{-1})
Set A [3] : based on corn stover dataset	
Glucose	4220 4706 5285 6697 7133 8529 8572
Xylose	5107 5648 6847 6855 7399 9386 9548
Galactose	4864 6697
Arabinose	4895
Mannose	4290 7094
Lignin	4409 4413 7110 7449 7453 8259 8321 8398 9413 9420
Ash	4957 6635 7133 7445 7449 9220 9251 9254 92798 9293 9297
Set B [4] : based on switchgrass dataset	
Glucose	4309 5235 5644 6828 8425
Xylose	5686 6766 7167 7229
Galactose	4313 5632 6750 9362
Arabinose	4305 5636 6897 9505 9520
Mannose	4497 5883 9578 9582 9605 9609 9621
Lignin	4297 5590 9478
Ash	4965 5339 5844 7422

1. Each row lists the chemical analyte and the significant variables for this analyte selected from 1558 spectral variables. All the values represent wavenumber.
2. Stepwise significance levels: inclusion level: 0.05; exclusion level: 0.10.
3. Set A presents the stepwise variable selection results conducted on the corn stover calibration dataset.
4. Set B presents the stepwise variable selection results conducted on the switchgrass calibration dataset.

the table, derived from the stepwise applied on corn stover calibration dataset. Theoretically, the selected variables (wavenumbers) indicated where the response chemical analyte had large absorbance; so the selections derived from both calibration datasets should have great resemblances. However, consistency was hardly observed, and this suggested that these two sets of stepwise selection results were data dependent thus not representative. Consequently, MLR developed upon the selection results will be subjective. Stepwise is only recommended when a very large dataset is available so that any influential points and noise are overwhelmed by real signal association. Another drawback of MLR is that it can only predict one chemical at a time. Compared with simultaneous predictions of seven chemical constituents, MLR takes more efforts and ignores the correlation between the chemical analytes. Therefore, MLR was not appropriate for this study.

Then the remaining comparison was between PLS and PCR. The cross-validation results of PCR and PLS applied to the individual corn stover and switchgrass dataset were shown in Table 14. For both calibration datasets (corn correlation and smaller RMSEP than PCR did throughout all the chemical

Table 14 The cross-validation results comparison between PCR and PLS [1]

		Glucose	Xylose	Galactose	Arabinose	Mannose	Lignin	Ash
Corn stover								
Correlation	PCR	0.916	0.787	0.556	0.767	0.780	0.765	0.666
	PLS	0.947	0.901	0.838	0.953	0.765	0.962	0.918
RMSEP [2]	PCR	1.756	1.910	0.413	0.717	0.321	2.628	1.275
	PLS	1.407	1.346	0.201	0.341	0.321	1.087	0.700
Switchgrass								
Correlation	PCR	0.972	0.936	0.831	0.875	0.648	0.901	0.953
	PLS	0.975	0.960	0.925	0.855	0.674	0.939	0.953
RMSEP	PCR	0.660	0.777	0.343	0.292	0.200	0.568	0.287
	PLS	0.633	0.620	0.235	0.374	0.203	0.458	0.266

1. Better model has correlation closer to 1 and lower RMSEP.
2. All the values for RMSEP have the unit of %w/w (dry-based).

stover and switchgrass, respectively), PLS regression method provided larger analytes. Therefore, PLS was better than PCR for biomass NIR modeling and thus was used as final modeling method in this study.

5.4 Predictive model results

5.4.1 Corn stover and switchgrass individual models

Based on all the selections discussed above, PLS regression were performed using the entire set of EMSC pretreated data in *log(1/R)* unit and all the chemicals were included in the modeling simultaneously. Leave-one-out cross-validation was utilized.

Figure 27 visually indicated good cross-validation results (correlation) of both individual models: the predicted chemical contents from the models were very close to the measured ones from wet chemistry analysis. The quantified cross-validation results are presented in Table 15. For switchgrass, the correlations between the measured values and the predicted values from the model were large (0.975, 0.960 and 0.939 %w/w, respectively) for the three major constituents, glucose, xylose, and lignin, which add up to over 80% weight percentage. RMSEP was 0.975 %w/w for glucose, 0.96 %w/w for xylose and 0.939 %w/w for lignin, and these values were small if compared to the original scale of each individual constituent: the prediction error rate (RMSEP divided by Mean) was only 2.5% for glucose, 3% for xylose, 2.4% for lignin. The correlation for mannose was poor (0.674) and RMSEP for the three minor monosaccharides had relatively larger error rate, approximately around 10% for galactose, arabinose, and up to 21% for mannose. The possible reasons are: 1) The weight percentage was small thus the contribution to the dataset variation was accordingly much less, while the composed latent factors during regression tended to explain from larger variance down. 2) Since the concentrations of these three monosaccharides in the hydrolyzate were very low, HPLC measurement of them was not as accurate as glucose and xylose.

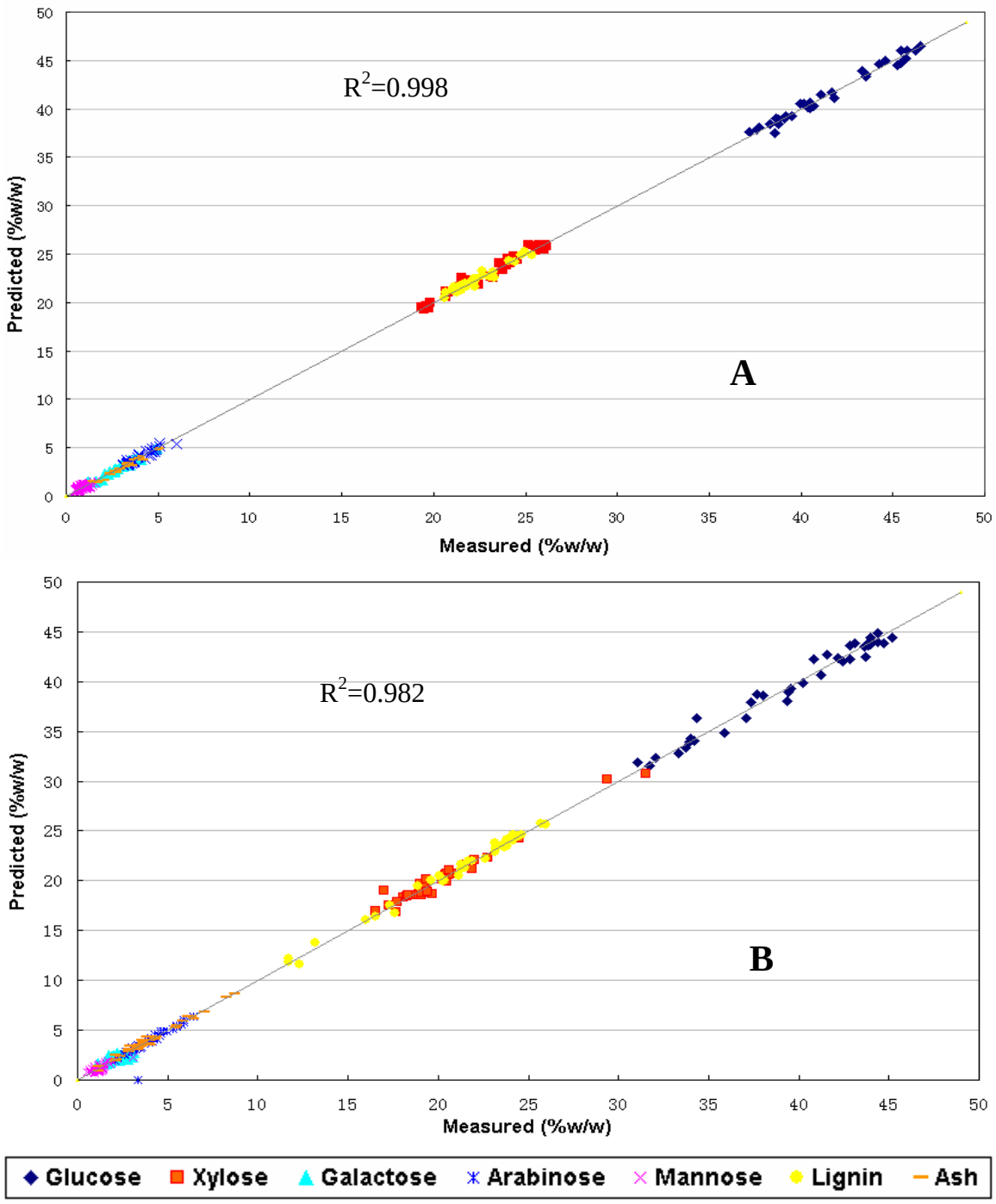


Figure 27 The predicted chemicals are plotted all together versus the measured chemical contents and R^2 is calculated based all the data points. A: switchgrass model; B: Corn stover model. Both graphs have the data points swarm around the linear diagonal line y=x.

Table 15 The cross-validation result of switchgrass individual model

Switchgrass	*Glucose*	*Xylose*	*Galactose*	*Arabinose*	*Mannose*	*Lignin*	*Ash*
Mean (%w/w)	41.37	23.04	2.65	4.05	0.94	22.28	2.95
Correlation	0.975	0.960	0.925	0.855	0.674	0.939	0.953
RMSEP (%w/w)	0.633	0.620	0.235	0.374	0.203	0.458	0.266
SEP (%w/w)	0.642	0.629	0.238	0.379	0.206	0.465	0.270
R*[1] *(%w/w)	9.33	6.85	2.01	2.90	1.02	4.75	3.83
***R/SEP*[2]**	14.533	10.89	8.45	7.65	4.95	10.215	14.185

1. The range of the dataset, calculated by the maximum minus the minimum

2. This parameter is used to evaluate the model performance target provided American Association of Cereal Chemists (AACC) standard :
   R/SEP≥4 calibration good for screening
   R/SEP≥10 calibration good for quality control
   R/SEP≥15 calibration good for research development

3) Like glucose and xylose, they are also carbohydrates, thus are very similar in terms of chemical bonds. So the corresponded spectra tended to have similar curvature, while PLS regression tended to give the credit of the variation to glucose and xylose for their overwhelming percentage weight. Although ash had a good correlation (close to 1), its error rate almost reached 10%. One possible cause was that inorganic compounds have no absorption in NIR region, while ash is primarily composed of inorganic compounds.

For corn stover (Table 16), likewise, the correlations between the measured values and the predicted values from the model were large: 0.947, 0.901 and 0.962 for the glucose, xylose, and lignin, respectively, RMSEPs were 0.975 %w/w for glucose, 0.96 %w/w for xylose and 0.939 %w/w for lignin, and error rate was 3.6% for glucose, 6.7% for xylose, 5.3% for lignin. With the reasons stated earlier, the RMSEP was still not very satisfying for galactose, arabinose, mannose and ash.

The prediction result visually seemed not as good as switchgrass model, however, the corn stover dataset covered a greater variability in their chemical composition. To make a fair comparison and judgment, a criteria published by the American Association of Cereal Chemist (AACC) was adopted, since there

Table 16 The cross-validation result of corn stover individual model

Corn Stover	*Glucose*	*Xylose*	*Galactose*	*Arabinose*	*Mannose*	*Lignin*	*Ash*
Mean (%w/w)	39.2	20.10	2.26	3.90	1.28	20.67	4.07
Correlation	0.947	0.901	0.838	0.953	0.765	0.962	0.918
RMSEP (%w/w)	1.407	1.346	0.201	0.341	0.321	1.087	0.700
SEP (%w/w)	1.427	1.359	0.204.	0.346	0.326	1.099	0.707
R (%w/w)	15.10	14.99	2.03	4.35	2.55	14.26	7.58
R/SEP	10.732	11.028	9.974	10.497	7.96	12.977	10.721

was no established standard for biomass NIR study. In the AACC Method 39-00 (1999), the recommended performance targets are: the model with *R/SEP*≧4 is qualified for screening calibration, ≧10 is acceptable for quality control, and ≧15 very good for research quantification. According to this standard, the performance levels of these two models turned out almost the same. For both corn stover and switchgrass, the models were suitable for the industrial quality control in terms of glucose, xylose, lignin and ash measurement, and were qualified for screening purposes in terms of galactose and mannose. And the only difference, in terms of performance targets, lied in the arabinose measurement: the switchgrass model of predicting galactose was at screening level, while the corn stover model was at quality control level.

5.4.2 One general model hypothesis investigation

5.4.2.1 Justification of the general model

First, PCA was conducted based on solely chemical information, and all PCs combinations failed to separate switchgrass and corn stover samples (Figure 28); more accurate description was the variation of corn stover chemical composition covered that of switchgrass dataset. Likewise, PCA results based on solely FT-NIR spectra of the corn stover and switchgrass samples again exhibited that corn stover and switchgrass samples overlap to a great degree

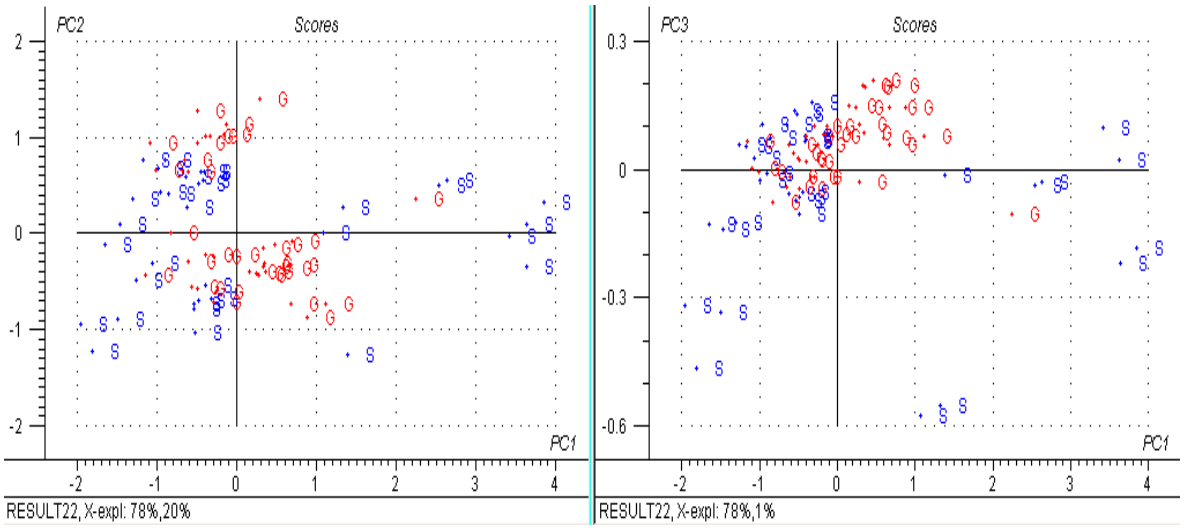


Figure 28 The PCA score plots based on the spectral data of corn stover and switchgrass data. PC1 vs. PC2. (left), PC1 vs. PC3 (right). S: corn stover; G: switchgrass

and are not differentiable (Figure 29). The rationale of developing one general model covering these two species was thus demonstrated.

5.4.2.2 Cross validation results

As it was reasonable to combine corn stover and switchgrass together for one general model, the general model of 71 observation (Section 4.3.2.1) was developed using the best modeling configurations (*log(1/R)*, EMSC, PLS) To make a fair comparison to the individual models, cross-validation was still used to evaluate the model accuracy (Table 17). Although the correlation decreased slightly compared to both of the two individual models, the prediction error rates (the ratio of RMSEP to mean) fell between them: 2.8%, 5.6%, and 6.3% for glucose, xylose, and lignin, respectively. According to AACC standard, these three major constituents still met the quality control criteria, and so did a minor constituent, ash. Also, the three minor saccharides were still good for screening.

Based on cross-validation results comparison, the general model for both corn stover and switchgrass achieved almost the same performance target as the two individual models. This general model was more robust without losing the accuracy.

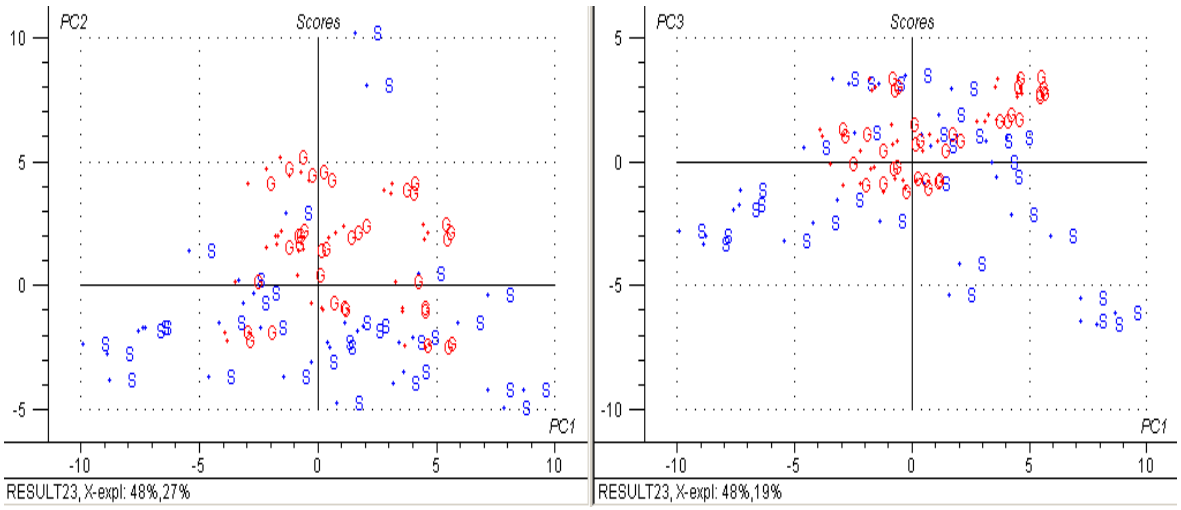


Figure 29 The PCA score plots based on the chemicals data of corn stover and switchgrass data. PC1 vs. PC2 (left), PC1 vs. PC3 (right). S: corn stover; G: switchgrass

Table 17 The cross-validation result of the general model

	Glucose	***Xylose***	***Galactose***	***Arabinose***	***Mannose***	***Lignin***	***Ash***
Descriptive Statistics of Calibration Dataset							
R [1] ***(%w/w)***	15.406	14.988	2.644	4.346	2.547	14.261	7.476
Mean (%w/w)	40.612	21.741	2.460	4.047	1.105	21.485	3.369
Model Performance Assessment							
Correlation	0.954	0.843	0.708	0.735	0.759	0.895	0.854
RMSEP (%w/w)	1.153	1.208	0.425	0.578	0.282	1.347	0.530
SEP (%w/w)	1.161	1.217	0.428	0.582	0.284	1.356	0.534
R/SEP	13.3 [2]	12.3 [2]	6.2 [3]	6.7 [3]	9.0 [3]	10.6 [2]	14.0 [2]

1. the data range calculated by Max-Min of the dataset
2. *R/SEP* greater than 10, thus good for quality control
3. *R/SEP* greater than 4, thus good for screening

5.4.2.3 Validation using independent data

The validation result based on the independent dataset (Section 4.3.2.3) is presented in Table 18. Compared to cross-validation results, RMSEP and SEP decreased for all the 7 constituents (Table 18). *R/SEP*s increased for all the constituents except for arabinose and lignin; and two constituents were improved to a higher criteria level than the cross-validation results: glucose was qualified for quantitative research while mannose was capable for quality control now. The results further validated that the developed general model can predict the chemical composition of corn stover and switchgrass accurately, and also proved that cross-validation is a conservative method to validate a developed NIR model.

Both cross-validation and independent validation showed promising predictive accuracy of the developed general model in examining switchgrass and corn stover.

5.4.2.4 The model prediction capability of wheat straw

As described in Section 4.3.2.3, the results of performing the five wheat straw samples are summarized in Table 19. The error rates were very low for the three major constituents, with only 1.56%, 2.03%, 2.90% for glucose, xylose, and lignin, respectively, and the error rates for arabinose and ash were under 10%.

Table 18 The results of validating the general model using independent dataset, including 5 corn stover and 5 switchgrass samples.

	Glucose	*Xylose*	*Galactose*	*Arabinose*	*Mannose*	*Lignin*	*Ash*
R [1] ***(%w/w)***	10.86	7.96	1.77	2.09	0.59	4.53	5.02
RMSEP (%w/w)	0.598	0.664	0.219	0.256	0.049	0.619	0.347
SEP (%w/w)	0.604	0.666	0.224	0.264	0.057	0.623	0.357
R/SEP	17.997 [2]	11.956 [3]	7.929 [4]	7.932 [4]	10.363 [3]	7.265 [4]	14.069 [3]

1. The data range of the 10 independent validation samples: 5 corn stover samples and 5 switchgrass samples (Table 5).
2. *R/SEP* was greater than 15, thus good for quantitative research.
3. *R/SEP* was greater than 10, thus good for quality control.
4. *R/SEP* was greater than 4, thus good for screening.

Table 19 The results of validating the general model using 5 wheat straw samples

	Glucose	*Xylose*	*Galactose*	*Arabinose*	*Mannose*	*Lignin*	*Ash*
Mean	42.269	20.253	2.036	3.143	0.749	24.050	2.358
R [1] ***(%w/w)***	2.30	2.29	0.58	0.70	0.13	2.87	0.84
RMSEP (%w/w)	0.659	0.412	0.311	0.201	0.135	0.698	0.213
SEP (%w/w)	0.662	0.414	0.316	0.205	0.136	0.703	0.216
Error Rate [2]	1.56%	2.03%	15.27%	6.39%	18.02%	2.90%	9.03%

1. the data range of the 5 wheat straw samples
2. Error rate: RMSEP divided by the mean of 5 wheat straw samples for each constituent respectively

The results showed the predictive potential of using the developed general model to predict wheat straw composition, especially in terms of the three major constituents. This test also proved that after spectral pretreatment FT-NIR probed only chemical characteristics. The results indicated great potential to use the developed general model to predict the chemical composition of wheat straw with accuracy, yet further validation investigation is favored by using wheat straw samples with larger variability.

5.4.3 HHV modeling

The cross-validation result (Figure 30) for the HHV model (Section 4.3.5) showed it was feasible to use FT-NIR to predict HHV. RMSEP was 53.23 J/g while the average HHV was 18932.1 J/g, so the prediction error rate was only 0.28%. Also, the correlation was 0.971 with the slope 0.964, indicating that the predicted HHV via the FT-NIR predicted model was very close to the measured HHV via the calorimeter. The modeling results showed the great potential to predict HHV via FT-NIR analyses approach.

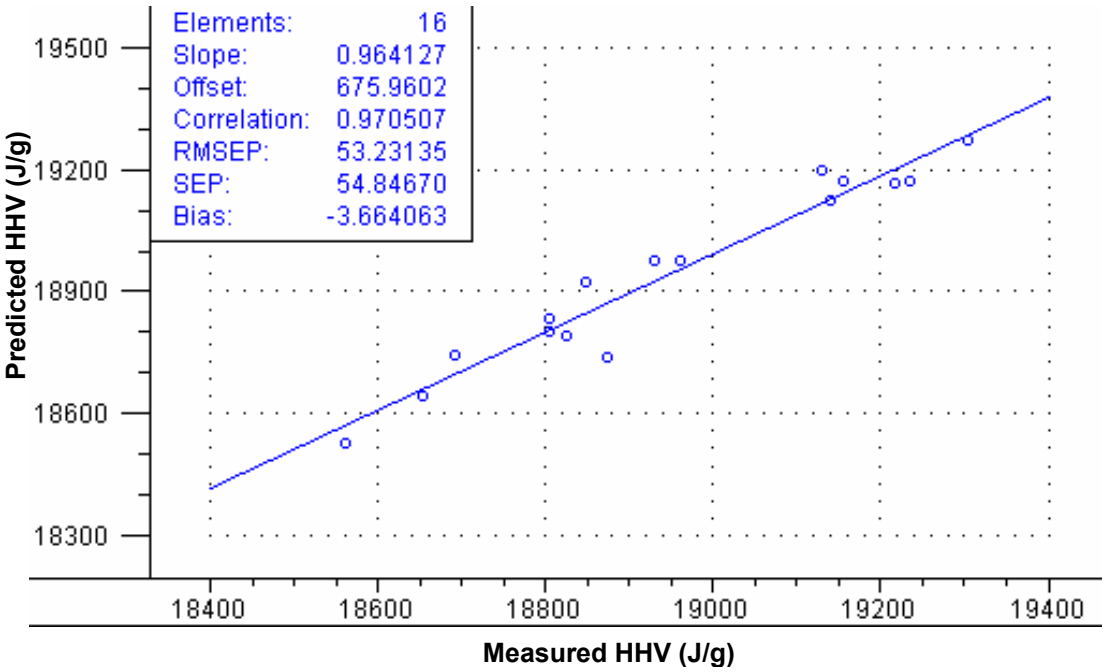


Figure 30 The plot of the predicted HHV vs. measured HHV with cross-validation results presented

CHAPTER VI CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Manual separation of botanic parts proved to be an effective approach to create variability for multivariate analysis. For corn stover, great variability exhibited amongst different botanic parts; and for switchgrass, different botanic parts provided variability greater than different cultivars did. Glucose, thus cellulose, was the major chemical constituents in the investigated biomass species, contributing to 30-50 % of the total dry weight. Both corn stover and switchgrass showed that internodal parts contained higher glucose content than nodes and leaves. Also, corn husks had the greatest total sugar content amongst all the botanic parts of corn stover, which suggested it as a very good feedstock for ethanol fermentation. Lignin and xylose had similar weight percentage in biomass and they were the other two major constituents in biomass, ranging roughly from 15% to 30%. The other three monosaccharides took up only small portion of biomass chemical composition, with descending sequence from arabinose (2~6%w/w), galactose (1~4%w/w), to mannose (0~2%w/w). Biomass leaves had higher lignin content compared to other botanic parts and also larger higher heating value as well. Generally, switchgrass exhibited less variation in chemical composition, compared to corn stover. Among switchgrass cultivars, Cave-in-rock and NC2-16 presented significantly higher glucose content than other cultivars, while NC1-16 had the lowest glucose content yet the top heating value.

Two NIR regions (4800 cm^{-1}~5450 cm^{-1} and 6804 cm^{-1}~7167 cm^{-1}) were correlated to the O-H bond attributed to water, and linear relationship was found between the spectral peak area and moisture content.

Studies proved that spectral pretreatment was necessary for the FT-NIR analysis on biomass samples and EMSC was the best pretreatment method to remove the physical interferences existing within biomass spectral sampling. Furthermore, this study concluded that it was better to present the spectral

intensity in *log(1/R)* rather than *K-M*; and partial least square was the best multivariate regression method for estimating the predictive FT-NIR model of biomass chemical composition . All these modeling configurations were recommended for the future NIR studies when biomass chemical composition is investigated.

The individual model developed specifically for switchgrass showed good accuracy: RMSEP of cross-validation was 0.633, 0.620, 0.235, 0.374, 0.293, 0.458 and 0.266 %w/w for glucose, xylose, galactose, arabinose, mannose, lignin and ash, respectively. RMSEPs were 1.427, 1.345, 0.201, 0.341, 0.321, 1.087, 0.700 %w/w for the corn stover individual model. The measurements of glucose, xylose, lignin and ash via both models were valid for industrial quality control, and the measurements for the other constituents were valid for industrial screening process.

Furthermore, this study justified, developed and validated a single predictive model of both corn stover and switchgrass. A combined model on both corn stover and switchgrass model was developed without losing much prediction accuracy. RMSEP of this general model via cross-validation was 1.153, 1.208, 0.425, 0.578, 0.282, 1.347, 0.530 %w/w for glucose, xylose, galactose, arabinose, mannose, lignin and ash, respectively, and RMSEPs via independent validation were even smaller. Using the general model to predict wheat straw resulted in small RMSEPs: 0.659, 0.412, 0.311, 0.201, 0.135, 0.698, 0.213 %w/w for the seven chemical constituents, which shows the potential of using this developed general model to predict wheat straw composition.

The study also showed that it was promising to apply FT-NIR techniques to predict HHV of biomass feedstock.

6.2 Future studies

The results showed that the switchgrass cultivar NC2-16 was rich in glucose, while NC1-16 had large heating value, which indicated their potential to provide higher yields in different productions. However, literature associated with these two cultivars, especially their yields, is rarely found. More research is

suggested to be conducted on these two switchgrass cultivars.

Internodal parts had higher glucose (or cellulose) content than nodes and leaves for both switchgrass and corn stover. Does this fact generally exist in all the biomass species? More biomass species need to be studied to verify this statement.

As for the FT-NIR modeling, there are several comments on future FT-NIR research. By inputting more samples with more variability in the chemical composition calibration model, the prediction accuracy will be further enhanced so that can be directly used for research purposes. More biomass species can be involved after the justification and hypothesis testing to make this method more robust and powerful. The compositional predictive models developed based upon the spectrometer utilized in this study can be transported to other NIR spectrometers via mathematical calibration and modeling. Efforts can be made to make the developed models more flexible and transferable. FT-NIR rapid analysis method shows great potential in HHV measurement, but sixteen samples are far from enough; more efforts are needed to fulfill the HHV calibration model.

.

LIST OF REFERENCES

LIST OF REFERENCES

AACC Method 39-00. 1999. Near-Infrared methods-Guidelines for model development and maintenance. *Approved Methods of the American Association of Cereal Chemists.* AACC Press, St. Paul, MN.

.

Albanell, E., J. Plaixats, A. Ferret, L. Bosc and F. Casanas. 1995. Evaluation of near-infrared reflectance spectroscopy for predicting stover quality trait in semi-exotic populations of maize. *Journal of the Science of Food and Agriculture*. 69(3): 269-273.

Alterthum, F and L. O. Ingram. 1989. Efficient ethanol production from glucose, lactose and xylose by recombinant Escherichia coli. *Appl Environ. Microbiol*. 55: 943–948.

ASTM E 1721-95. 1995. Standard test method for determination of acid-insoluble residue in biomass. *Annual Book of ASTM Standards*, 11.05. West Conshocken, PA: ASTM International.

ASTM E 1758-95. 1995. Determination of carbohydrates in biomass by high performance liquid chromatography. *Annual Book of ASTM Standards*, 11.05. West Conshocken, PA: ASTM International.

ASTM E 1755-01. 2002. Standard Test Method for Ash in Biomass. *Annual Book of ASTM Standards*, 11.05. West Conshocken, PA: ASTM International.

Baianu, I.C., T. You, D.M. Costescu, P.R. Lozano, V. Prisecaru and R.L. Nelson. 2004. High-resolution nuclear magnetic resonance and near-infrared determination of soybean oil, protein, and amino acid residues in soybean seeds. *Oil Extraction and Analysis: Critical Issues and Comparative Studies*. 94th AOCS Annual Meeting & Expo, Kansas City, MO, United States. 193-240.

Bajcsy, R., S.W. Lee and A. Leonardis. 1996. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *International Journal of Computer Vision*. 17(3):241-272.

Balleirini dos Santos, R., L. Chagas de Sousa and J.L. Gomide. 2006. Use of near-infrared (NIR) spectroscopy in evaluation of wood chips. *Papel.* 67(5): 84-93.

Baxter, L.L. 1993. Ash deposition during biomass and coal combustion: a mechanistic approach. *Biomass Bioenergy*. 4(2):85–102.

Bell, R.J. 1972. Introductory Fourier Transform Spectroscopy, Academic, New York.

Boateng, A.A, K.B. Hicks and K.P. Vogel. 2006. Pyrolysis of switchgrass (Panicum virgatum) harvested at several stages of maturity. *Journal of Analytical and Applied Pyrolysis*. 75:55-64.

Bothast R. J., B.C. Saha, A. V. Flosenzier and L. O. Ingram. 1994. Fermentation ofl-arabinose,d-xylose andd-glucose by ethanologenic recombinantKlebsiella oxytoca strain P2. *Biotechnol Lett* 16: 401–406.

Brown, W.F., J.E. Moore, W.F. Kunkle, C.G. Chambliss and K.M. Portier. 1990. Forage testing using near infrared reflectance spectroscopy. *Journal of Animal Science*. 68: 1416-1427.

Christian D.G. and H.W. Elbersen. 1998. Switchgrass (Panicum virgatum L.). In: N. El Bassam. Energy plant species. Their use and impact on environment and development. London: James and James publishers 257-263.

Church, J.S., J.A. O'Neill and A.L. Woodhead. 1999. A comparison of vibrational spectroscopic methods for analyzing wool/polyester textile blends. *Textile Research Journal*. 69(9):676-684.

Confalonieri, M., G. Lombardi, M. Bassignana and M. Odoardi. 2004. Analysis of quality constituents of natural alpine swards with near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy.* 12(6):411-417.

Cozzolino, D., A. Fassio, E. Fernandez, E. Restaino and A. La Manna. 2006. Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy. *Animal Feed Science and Technology*. 129(3-4): 329-336.

Cozzolino, D., A. Fassio and A. Gimenez. 2000. The use of near-infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants. *Journal of the Science of Food and Agriculture*. 81(1):142-146.

Czarnik-Matusewicz, B., K. Murayama, R. Tsenkova and Y. Ozaki. 1999. Analysis of near-infrared spectra of complicated biological fluids by two-dimensional correlation spectroscopy: protein and fat concentration-dependent spectral changes of milk. *Applied Spectroscopy.* 53(12):1582-1594.

Dahm, D.J. and K.D. Dahm. 1995. Letter: math pretreatment of NIR reflectance

data: *log(1/R)* vs F(R). *Journal of Near Infrared Spectroscopy.* 3(1): 53-59.

de Aldana, B.R. Vazquez, B. and Garcia-Criado, A. Garcia-Ciudad and M.E. Perez-Corona. 1996. Non-destructive method for determining ash content in pasture samples: application of near infrared reflectance spectroscopy. *Communications in Soil Science and Plant Analysis*. 27(3 & 4):795-802.

De Noord, O.E. 1994. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models, *Chemom. Intell. Lab. Syst...* 25: 85.

Demirba, A. 2001. Biomass resource facilities and biomass conversion processing for fuels and chemicals. *Energy Conversion and Management*. 42(11):1357-1378.

Demirbas, A. 2005. Heavy metal contents of fly ashes from selected biomass samples. *Energy Sources*. 27:1269–1276.

DOE 2004. Understanding Biomass as a Source of Sugars and Energy. Available at: http://www1.eere.energy.gov/biomass/understanding_biomass.html. accessed on 14. December, 2006.

Donnelly, E. D. and J. I. Wear. 1972. Acid detergent method for reduction of tannin interference in determining lignin of sericea lespedeza. *Agronomy Journal*. 64(6): 838-839.

Dou, Y., H. Mi, L. Zhao, Y. Ren and Y. Ren. 2006. Radial basis function neural networks in non-destructive determination of compound aspirin tablets on NIR spectroscopy. *Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy*. 65A(1): 79-83.

Duffy, M.D. and V.Y. Nanhou. 2002. Costs of producing switchgrass for biomass in Southern Iowa. *Trends in new crops and new users*. 267-275.

Edwards, C. S. 1973. Determination of lignin and cellulose in forages by extraction with triethylene glycol. *Journal of the Science of Food and Agricultur*e. 24(4): 381-388.

Elbersen, H.W., D.G. Christian, W. Bacher, E. Alexopoulou, V. Pignatelli and D. van den Berg. 1998. Switchgrass variety choice in Europe. *Final Report FAIR 5-CT97-3701 "Switchgrass"*. 35-41.

Esbensen, K.H. 2004. *Multivariate Data Analysis in practice*. 5th Edition. CAMO process AS, Oslo, Norway..

Fardim, P., M. Ferreira, and N. Duran. 2002. Multivariate calibration for quantitative analysis of eucalypt kraft pulp by NIR spectrometry. *Journal of Wood Chemistry and Technology*. 22(1): 67-81.

Fardim, P., M. Ferreira, N. Duran. 2005. Determination of mechanical and optical properties of eucalyptus kraft pulp by NIR spectrometry and multivariate calibration. *Journal of Wood Chemistry and Technology*. 25(4): 267-279.

Fearn, T.. 2000. Are two pretreatment better than one? *NIR news*. 11(6):14.

Fennema, O.R. 1996. *Food Chemistry*. (3rd Edition). Marcel Dekker, Inc. 205-206.

Flinn, P.C., R.G. Black, L. Iyer, J.B. Brouwer and C. Meares. 1998. Estimating the food processing characteristics of pulses by near infrared spectroscopy, using ground or whole samples. *Journal of Near Infrared Spectroscopy*. 6(1-4): 213-220.

Flinn, P.C. and A.B. Blakeney. 1996. Color measurement of hay using near infrared spectroscopy. *Near Infrared Spectroscopy: The Future Waves*, Proceedings of the International Conference on Near Infrared Spectroscopy, 7th, Montreal, Aug. 6-11, 1995 Meeting Date 1995, 542-544.

Freitas, M.P, A. Sabadin, L.M. Silva, F.M, Giannotti do Couto Debora A, E. Tonhi, R.S. Medeiros, G.L. Coco, V.F.T. Russo and J.A. Martins. 2005. Prediction of drug dissolution profiles from tablets using NIR diffuse reflectance spectroscopy: a rapid and nondestructive method. *Journal Of Pharmaceutical And Biomedical Analysis*. 39(1-2):17-21.

Garcia-Ciudad, A., B. Garcia-Criado, M.E. Perez-Corona, B.R. Vazquez De Aldama and A.M. Ruano-Ramos. 1993. Application of near - infrared reflectance spectroscopy to chemical analysis of heterogeneous and botanically complex grassland samples. *Journal of the Science of Food and Agriculture*. 63(4): 419-426.

Gnandesikan, R. 1977. *Method for Statistical Data Analysis of Multivariate Observations*. John Wiley and Sons, New York. 306.

Goodchild, A. V., F.J. El Haramein, A. Abd El Moneim, H.P.S. Makkar and P.C. Williams. 1998. Prediction of phenolics and tannins in forage legumes by near infrared reflectance. *Journal of Near Infrared Spectroscopy*. 6(1-4): 175-181.

Griffiths, P.R. and J.A. de Haseth. 1986. *Fourier Transform Infrared spectrometry*. John Wiley & Sons, New York.

Hames, B.R., S.R. Thomas, A.D. Sluiter, C.J. Roth and D.W. Templeton. 2003. Rapid biomass analysis: New tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production. *Applied Biochemistry and Biotechnology.* 105-108.

Higuchi, T. 1990. Lignin biochemistry: Biosynthesis and biodegradation. *Wood Sci*. *Technology.* 24: 23-63.

Hildrum, K.I, T. Isaksson, T. Naes, M. Rodbotten and P. Lea. 1996. Near infrared reflectance spectroscopy in the prediction of sensory properties of beef. *Journal of Near Infrared Spectroscopy*. 3(2): 81-87.

Hoel, M. and S. Kverndokk. 1996. Depletion of fossil fuels and the impacts of global warming. *Resource and Energy Economics*. 18(2): 115-136.

Hong J.H., K. Ikeda, I. Kref and K. Yasumoto. 1996. Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch, amylose, and tannin in buckwheat flours. *Journal of Nutritional Science and Vitaminology.* 42(4): 359-66.

Hopkins, A.A., K.P. Vogel, K.J. Moore, K.D. Johnson and I.T. Carlson. 1995. Genotypic variability and genotype × environment interactions among switchgrass accessions from the midwestern USA. *Crop Science*. 35(2): 565-571.

Hu, S.B., A. Lillquist, M.A. Arnold and J.M. Wiencek. 2000. Partial least square analysis of lysozyme near-infrared spectra. *Applied Biochemistry and Biotechnology*. 87(3):153-163.

Hu, T. Q.. 2002. Chemical modification, properties, and usage of lignin. Kluwer Academic/Plenum Publishers. 81-82.

Isaksson, T. and T. Naes. 1988. The Effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy*. 42(7): 1273-1284.

Isaksson, T., G. Toegersen, A. Iversen and K.I. Hildrum. 1995. Non-destructive determination of fat, moisture and protein in salmon fillets by use of near-infrared diffuse spectroscopy. *Journal of the Science of Food and Agriculture*. 69(1): 95-100.

Johnson, D.E. 1998. *Applied Multivariate Methods for Data Analysis*. Duxbury Press. Chapter 7:246-247.

Jung, H. G. 1989. Forage lignins and their effects on fiber digestibility. *Agronomy Journal.* 81(1): 33-38.

Kelley, S.S., T.G. Rials, R. Snell, L.H. Groom and A. Sluiter. 2004. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Science and Technology*. 38(4): 257-276.

Klasek, S.E.. 2006. *Terminal velocity determinations for component separation of biomass*. MS thesis. University of Tennessee, Dept. of Biosystems Engineering & Soil Science.

Koenig, J. L.. 2001. *Infrared and Raman Spectroscopy of Polymers*. Published by Rapra Technology. 7-8

Kortum, R.. 1969. R*eflectance Spectroscopy*. Springer-Verlag, New York. 103-106.

Kramer R.. 1998. *Chemometic Techniques for Quantitative Analysis*. Marcel Dekker, Inc.

Kracl, N. P. 1986. Corn harvester combine assembly with corn removal attachment. US Patent 4,583,354.

Kucuk, M.M. and A. Demirbas. 1997. *Biomass conversion processes*. *Fuel and Energy Abstracts*. 38(3): 163-163(1).

Lemus, R., E.C. Brummer, K.J. Moore, N.E. Molstad, C.L. Burrasb and M.F. Barkerb. 2002. Biomass yield and quality of 20 switchgrass populations in southern Iowa. *Biomass and Bioenergy*. 23: 433-442.

Lestander, T. A. and C. Rhen. 2005. Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst* (*Cambridge, United Kingdom)*. 130(8):1182-1189.

Liu, Y., S. Kokot and T.J. Sambi. 1998. Vibrational spectroscopy investigation of Australian cotton cellulose fibers. Part 2. A Fourier transform near-infrared preliminary study. *Analyst* (Cambridge, United Kingdom) 123(8): 1725-1728.

Liu, Y. and Y. Ying. 2005. Use of FT- NIR spectrometry in non-invasive measurements of internal quality of Fuji' apples. *Postharvest Biology and Technology*. 37(1): 65-71.

Liu, Z., 1996. A rapid determination of maize powder mixed in soybean meal by

near infrared spectroscopy. *Fenxi Shiyanshi.* 15(5):84-88.

Maertens, K., P. Reyns and J. Baerdemaeker. 2004. On-line measurement of grain quality with NIR technology. *Transactions of the ASAE*. 47(4): 1135-1140.

Manley, M., L. Van Zyl and B.G. Osborne. 2002. Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein, and moisture content of whole wheat flour. *Journal of Near Infrared Spectroscopy*. 10(1):71-76.

Mark, H. and J. Workman. 1987. The Effect of Principle Component Analysis in Regard to the Noise Sensitivity of calibration Equations. Paper #030, Pittsburgh Conference and Exposition, Atlantic City, N. J. March 9-13, 1987.

Marten, G.C., G.E. Brink, D.R. Buxton, J.L. Halgerson and J. S. Hornstein. 1984. Near infrared reflectance spectroscopy analysis of forage quality in four legume species. *Crop Science.* 24: 1179-1182.

Martens, H. and T. Naes, 1987. Multivariate calibration by data compression. *Near-infrared Technology in the Agricultural and Food Industries*. by Williams, P. and K. Norris (eds), AASS, St. Paul, MN. 57-87.

Martens, H., J.P. Nielsen and S.B. Engelsen. 2003. Light scattering and light absorbance separated by extended multiplicative signal correction. *Application to Near-Infrared Transmission Analysis of Powder Mixtures. Anal. Chem*. 75: 394-404.

Martens, H. and E. Stark. 1991. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical & Biomedical Analysis*. 9(8): 625-635.

McCarthy, W.J. and G.J. Kemeny. 2001. Fourier Transform Spectraphotometers in Near-infrared. *Handbook of Near-infrared Analysis. (2nd Ed.)*: 71-88.

McLaughlin, S., J. Bouton, D. Bransby, B. Conger, W. Ocumpaugh, D. Parrish, C. Taliaferro, K. Vogel, and S. Wullschleger. 1999. *Developing Switchgrass as a Bioenergy Crop*. Reprinted from: Perspectives on new crops and new uses. J. Janick (ed.), ASHS Press, Alexandria, VA. 282-299.

Melchhinger, A.E., G.A. Schmidt and H.H. Geiger. Evaluation of Near Infra-red Reflectance Spectroscopy for Predicting Grain and Stover Quality Traits in

Maize. *Plant Breeding*. 97(1): 20-29.

Michell, A.J. and L.R. Schimleck. 1996. NIR spectroscopy of woods from Eucalyptus globules. *Appita Journal.* 49: 23–26.

Montross, M.D. and C.L. Crofcheck. 2004. Effect of stover fraction and storage method on glucose production during enzymatic hydrolysis. *Bioresource Technology*. 92: 269–274.

Munck, L. 2006. Conceptual validation of self-organisation studied by spectroscopy in an endosperm gene model as a data-driven logistic strategy in chemometrics. *Chemometrics and Intelligent Laboratory Systems.* 84(1-2): 26-32.

Nikolich, K., C. Sergides and A. Pittas. 2001. The application of near infrared reflectance spectroscopy (NIRS) for the quantitative analysis of hydrocortisone in primary materials. *Journal of the Serbian Chemical Society*. 66(3):189-198.

Nousiainen, J., S. Ahvenjarvi, M. Rinne, M. Hellamaki, and P. Huhtanen. 2004. Prediction of indigestible cell wall fraction of grass silage by near infrared reflectance spectroscopy. *Animal Feed Science and Technology*. 115(3-4): 295-311.

Osborne, B.G..1981. Principles and practice of near infrared (NIR) reflectance analysis. *International Journal of Food Science & Technology*. 16(1):13-19.

Osborne, B.G. 1986. *Near Infrared Spectroscopy in Food Analysis.* BRI Australia Ltd, North Ryde, Australia.1-14.

Pasikatan, M.C., J.L. Steele, C.K. Spillman and E. Haque. 2001. Near-infrared reflectance spectroscopy for on-line particle size analysis of powders and ground materials. *J. Near Infrared Spectroscopy*. 9: 153-164.

Patzek, T.W. 2004. Thermodynamics of the corn-ethanol biofuel cycle. *Critical Reviews in Plant Sciences.* 23(6): 519-567.

Peirs, A., N. Scheerlinck, K. Touchant and B.M. Nicolai. 2002. Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems Engineering*. 81(3):305-311.

Poke, F. S. and C. A. Raymond. 2006. Predicting extractives, lignin, and cellulose contents using near infrared spectroscopy on solid wood in eucalyptus globulus. *Journal of Wood Chemistry & Technology*. 26(2): 187-199.

Rantanen, J., E. Rasanen, J. Tenhunen, M. Kansakoski, J-P. Mannermaa and J. Yliruusi. 2000. In-line moisture measurement during granulation with a four-wavelength near infrared sensor: an evaluation of particle size and binder effects. *European Journal of Pharmaceutics and Biopharmaceutics*. 50(2): 271-276.

Raveendran, K. and A. Ganesh. 1996. Heating value of biomass and bimass pyroplysis products. *Fuel*. 75(15):1715-1720.

Rodriguez-Saona, L. E., F. S. Fry, M. A. McLaughlin and E. M. Calvey. 2001. Rapid analysis of sugars in fruit juices by FT- NIR spectroscopy. *Carbohydrate research.* 336(1): 63-74.

Saiz-Abajo, M. J., B. H. Mevik, V. H. Segtnan, and T. Naes. 2005. Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. *Analytica Chimica Acta*. 533(2):147-159.

Sakakibara, A. 1983. Chemical structure of lignin related mainly to degradation products. *Recent Advances in Lignin Biodegradation Research*. Tokyo: UNI Publisher. 12-33

Sanderson, M. A.; F. Agblevor, M. Collins, D. K. Johnson. 1996. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. *Liquid Fuel and Industrial Products from Renewable Resources,* Proceedings of the Liquid Fuel Conference, 3rd, Nashville, Sept. 15 -17. 1996. 37-42.

Sato, H., M. Shimoyama, T. Kamiya, T. Amari, S. Sasic, T. Ninomiya, H.W. Siesler and Y. Ozaki. 2003. Near infrared spectra of pellets and thin films of high-density, low-density and linear low-density polyethylenes and prediction of their physical properties by multivariate data analysis. *Journal of Near Infrared Spectroscopy*. 11(4): 309-321.

Savitzky A. and M.J.E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36: 1627-1639.

Schimleck, L. R., C. A. Raymond, C. L. Beadle, G. M. Downes, P. D. Kube and J. French. 2000. Applications of NIR spectroscopy to forest research. *Appita journal*. 53(6): 458-464.

Schonkopf, S., H. Martens and B. Alsberg. 1992. EMSC and SIS-multivariate preprocessing of NIR spectra by utilizing spectral background knowledge. Making Light Work: Adv. Near Infrared Spectroscopy, Int. Conf. *Near Infrared Spectrosc.* 185-90.

Schwanninger, M., B. Hinterstoisser, N. Gierlinger, R. Wimmer and J. Hanger.

2004. Application of Fourier Transform near infrared spectroscopy (FT-NIR) to thermally modified wood. *Holz als Roh- und Werkstoff.* 62(6): 483-485.

Shapouri, H. and P. Gallagher. 2005. USDA's 2002 ethanol cost-of-production survey. *Agricultural Economic Report.* No.841.

Shimoyama, M., K. Matsukawa, H. Inoue, T. Ninomiya and Y. Ozaki. 1999. Non-destructive analysis of photo-degradation of poly (methyl methacrylate) by near infrared light-fiber spectroscopy and chemometrics. *Journal of Near Infrared Spectroscopy*. 7(1): 27-32.

Sluiter, Amie , B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton. 2006. Determination of structural carbohydrates and lignin in biomass. *NREL BAT Team Laboratory Analytical Procedure*.

Smith, K. F. and P. C. Flinn. 1991. Monitoring the performance of a broad-based calibration for measuring the nutritive value of two independent populations of pasture using near infrared reflectance (NIR) spectroscopy. *Australian Journal of Experimental Agriculture*. 31: 205-210.

Sokhansanj, S., A. Turhollow, J. Cushman and J. Cundiff. 2002. Engineering aspects of collecting corn stover for bioenergy. *Biomass and Bioenergy.* 23(5): 347-355.

Steyerberg, E. W., M. J. Eijkemans, F. E. Jr. Harrell, J. D. Habbema. 2000. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 19:1059-1079.

Sun, J. 1996. A multivariate principal component regression analysis of NIR data. *Journal of Chemometrics*. 10(1): 1-9.

Sverzut, C. B., L. R. Verma and A. D. French. 1987. Sugarcane analysis using near infrared spectroscopy. *Transactions of the ASAE* 30(1): 255-258.

Sánchez, F. Cuesta, J. Toft, B. van den Bogaert, D.L. Massart, S.S. Dive and P. Hailey. 2004. Monitoring powder blending by NIR spectroscopy. *Fresenius' Journal of Analytical Chemistry*. 352(7-8): 771-778.

Thermo application note TN-00128. 2002. FT-IR vs. Dispersive Infrared. Theory of Infrared Spectroscopy Instrumentation. Thermo Nicolet Corporation.

Tosi, S., M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti, and D. Matteuzzi. 2003.

Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. *Biotechnology progress*. 19(6): 1816-1821.

Towne G. and C. Owensby. 1984. Long-term effects of annual burning at different dates in ungrazed Kansas tallgrass prairie. *Jounal of Range Management*. 37(5): 392-397.

Van Zyl, C, B. A. Prior and J. C. Du Preez. 1988. Production of ethanol from sugar cane bagasse hemicellulose hydrolysate by Pichia stipitis. *Applied Biochemistry and Biotechnology*. 17(11): 357-370.

Veraverbeke, E.A., J. Lammertyn, B.M. Nicolaie and J. Irudayaraj. 2005. Spectroscopic Evaluation of the Surface Quality of Apple. *Journal of Agricultural and Food Chemistry.* 53(4): 1046-1051.

Wei, L., H. Jiang, J. Li, Y. Yan and J. Dai. 2005. Predicting the chemical composition of intact kernels in maize hybrids by near infrared reflectance spectroscopy. *Guangpuxue Yu Guangpu Fenxi*. 25(9):1404-1407.

Winegartner, E.C. (Ed.) 1974. *Coal fouling and slagging parameters*, American Society of Mechanical Engineers, New York.

Wold S. and M. Sjostrom. 1998. Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems*. 44(1): 3-14.

Workman, J. and J. Brown. 1996. A new standard practice for multivariate, quantitative infrared analysis, Part 1. *Spectroscopy*. 11(2): 48-51.

Workman, J. and J. Brown. 1996. A new standard practice for multivariate, quantitative infrared analysis, Part 2. *Spectroscopy*. 11(9): 24-30.

Workman, J. J.. 1999. Review of process and non-invasive near-infrared and infrared spectroscopy: 1993–1999. *Applied Spectroscopy Reviews.* 34(1&2): 1-89.

Xiccato, G., A. Trocino, A. Carazzolo, M. Meurens, L. Maertens and R. Carabano. 1999. Nutritive evaluation and ingredient prediction of compound feeds for rabbits by near-infrared reflectance spectroscopy (NIRS). *Animal Feed Science and Technology*. 77:201-212.

Zimmer, E., P. A. Gurrath, Chr. Paul, B.S. Dhillon, W.G. Pollmer and D. Klein. 1990. Near infrared reflectance spectroscopy analysis of digestibility traits of maize stover. *Euphytica*. 48(1): 73-81.

Znidarsic, T., J. Verbic and D. Babnik. 2005. Prediction of chemical composition and energy value of hay by near-infrared reflectance spectroscopy (NIRS). *Acta Agriculturae Slovenica*. 86(1): 17-25.

APPENDICES

APPENDIX A
NOMECLATURE

Nomenclature

g	Grams	min	minutes
mg	milligrams	sec	seconds
ml	milliliters	hr	hours
mm	millimeters	J/g	joule per gram
ºC	degrees Celsius	cm^{-1}	wavenumber
%w/w	weight percentage (dry base)	nm	nanometer

Abbreviation

NIR	near infrared
FT	Fourier transform
μ	mean
σ	standard error
α	significant level
PC	principle component
PCA	principle component analysis
PCR	principle component regression
PLS	partial least squares regression
MLR	multiple linear regression
HCA	hierarchical clustering analysis
K-M	Kubelka-Munk
MC	moisture content
HHV	higher heating value
MSC	multiplicative scatter correction
EMSC	extended multiplicative signal correction
SNV	standard normal variate
RMSEP	root mean square error for prediction
SEP	standard error for prediction
R/SEP	ratio of the data range to standard error for prediction

Hu	Husk (DK64-10RR)
Hs	Husk (Incredible)
G	Switchgrass straw
A	Alamo
K	Kanlow
InterN	Internode
1^{st} +SNV	1^{st} derivative+ SNV
SNV+1^{st}	SNV+ 1^{st} derivative
2^{nd} +SNV	2^{nd} derivative+ SNV
SNV+2^{nd}	SNV+ 2^{nd} derivative

APPENDIX B
EXPERIMENTAL UTILITIES

List of software and devices used in the experiment and data processing

Software	***Major Usage***
The Unscrambler 9.2 (CAMO)	Pretreatments, PCA, PCR, PLS
NCSS 2004 (NCSS)	Stepwise, HCA, descriptive statistics.
Excel 2003 (Microsoft)	Other simple computation and graphs
Varian Resolution Pro (Varian Inc.)	Spectrometer control
GC Chemstation (Agilent Tech.)	HPLC monitor and integration
UV Probe 2.00 (Shimadzu Co.)	UV-Vis control
Hardware	***Major Usage***
Wiley Mini Mill (Thomas Scientific)	Sample preparation.
Varian FT-IR spectrometer Excalibur 3100	NIR spectra
Pike NIR IntegratIR Integrating Sphere Accessory	NIR spectra
BioRad Aminex HPX-87P column (300 x 7.8 mm)	Sugar detection
BioRad H+/CO3 deashing guard column	Sugar detection
Waters 410 refractive index detector	Sugar detection
SSI Lab Alliance Series I HPLC Pump	Sugar detection
Accumet Basic pH meter AB15	Neutralization pH control
Thelco convection oven	Drying
Barnstead Termolyne 1300 Furnace	Ash and AIL measurement
Fisher Scientific Isotemp 3006 (Water bath)	Hydrolysis
Schimatzu UV-1700 UV-spectrophotometer	ASL detection
Mass Balance (accurate to 0.1mg)	Weighing
Autoclave	Hydrolysis
IKA calorimeter system C 200	HHV
Carver 4350.L pelletizer	Sample preparation for HHV
Olympus SZH10 Research Stereo (microscope)	Magnification
Sony MAVICA MVC-CD500	Imaging

APPENDIX C
STATISTICS CALCULATIONS

Range (*R*)

$$R = \max(x_i) - \min(x_i)$$

Standard error (σ)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$$ where: $\overline{x}$ is mean

Euclidean distance (d_{jk})

Assume a dataset has m observations, each of which contains n variables. In terms of matrices, this dataset can be expressed as a m×n matrix with observations as rows and variables as columns.

j and k are two rows (observations), and the Euclidean distance is calculated as:

$$d_{jk} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - x_{ik})^2}{n}}$$

Mahalanobis distances (d_{jk}')

With the same assumption as above, the Mahalanobis distance is calculated as:

$$d_{jk}' = \frac{\sum_{i=1}^{n}\left|x_{ij} - x_{ik}\right|}{n}$$

Covariance (between X and Y) ($Cov(x, y)$)

$$Cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

where: $\overline{x}$, $\overline{y}$ are the means of X and Y, respectively

Correlation (between X and Y) (*r*)

$$r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

where $Cov(x, y)$ is the covariance between X and Y

σ represents standard error.

Root Mean Square Error for Prediction (*RMSEP*)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$$

Where: $\hat{y}_i$ is the predicted value for the i^{th} observation

y_i is the measured of value of i^{th} observation

n is the total number of observations

Bias: The mean of the regression errors

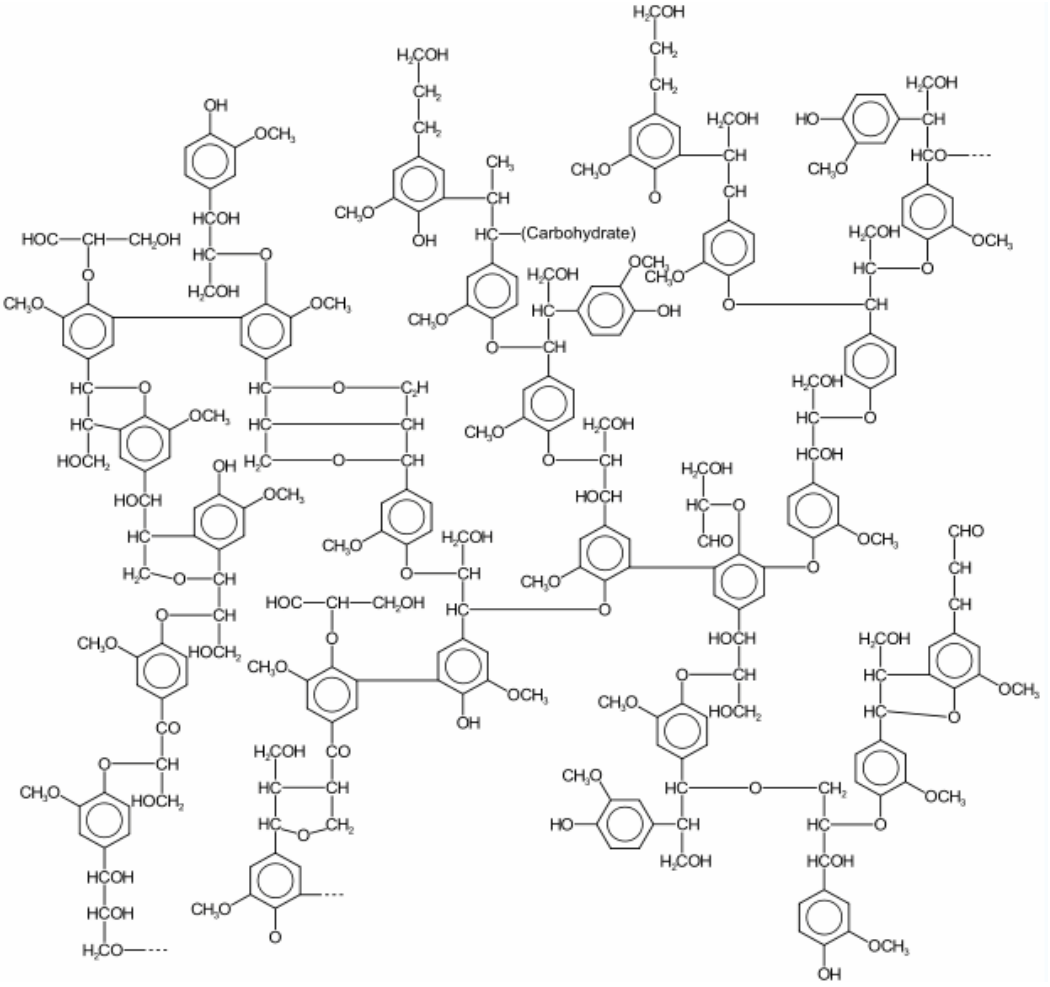
$$Bias = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Standard Error for Prediction (*SEP*)

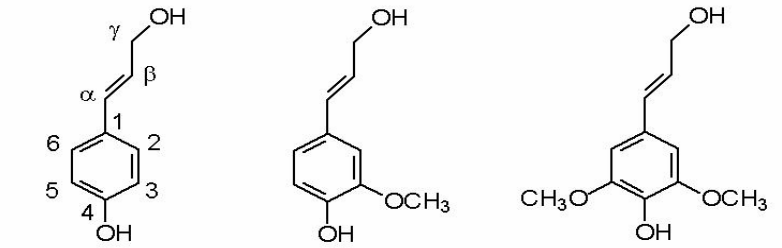
$$SEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i - Bias)^2}{n-1}}$$

APPENDIX D
LIGNIN STRUCTURAL INFOMATION

An example of a possible lignin structure



Structural model of softwood lignin (Sakakibara, 1983)



The three common monolignols in lignin

VITA

Lu Liu (Shirley) was born in Fuzhou, Fujian, China. She lived in her hometown until admitted by Tongji University in 2000. She went to Shanghai and finished her Bachelor's degree in Applied Chemistry. After graduation in 2004, she took an internship position at the Institute of Biotechnology in Fuzhou University and began preparing for the application of overseas study. In the summer of 2005, she flew over the Pacific to the United States and started her academic career in Biosystems Engineering at the University of Tennessee, Knoxville. Two years later, she completed a Master's degree in Biosystems Engineering and immediately began working towards a Ph.D in the same field.